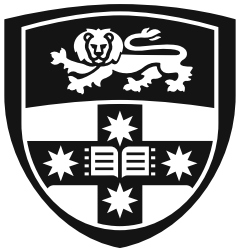
The Unrealised Potential of AI Solutions for Pasture-based Dairy Systems

Blessing Nnenna Azubuike

A thesis submitted in the fulfilment of the requirements for the degree of Doctor of Philosophy

School of Life and Environmental Sciences
Faculty of Science
The University of Sydney

2026



THE UNIVERSITY OF SYDNEY

PREFACE

This thesis is presented in the format of a ‘thesis with publications’ and is written in Australian English. Several chapters have been prepared as standalone manuscripts, some of which have been published or are under peer review. Hyperlinks have been enabled throughout the digital version for ease of navigation. All studies conducted as part of this thesis are my own original work. External contributions and assistance have been acknowledged where appropriate. The research was undertaken as part of the DairyUP project, focusing on precision nutrition, pasture biomass estimation, grazing detection, and data-driven decision support for pasture-based dairy systems.

Blessing Nnenna Azubuike

December 30, 2025

STATEMENT OF ORIGINALITY

This thesis is submitted to The University of Sydney in fulfilment of the requirements for the degree of Doctor of Philosophy. I certify that the intellectual content of this thesis is the product of my own original work, except where due acknowledgement is made. This thesis has not been submitted, in whole or in part, for any other degree or qualification at any university or institution. All external sources, datasets, collaborative inputs, and assistance received during this research have been properly acknowledged.

Blessing Nnenna Azubuike

December 30, 2025

AUTHORSHIP ATTRIBUTION STATEMENT

This thesis contains six research chapters which are the results of my own investigations.

Chapter 1 presents a general introduction and research aims.

Chapter 2 presents a review of published literature under consideration with *Computers and Electronics in Agriculture* and is titled 'Artificial intelligence in pasture-based dairy systems: Advances in precision feeding, remote sensing, and grazing management'. Blessing Nnenna Azubuike (B.N.A) conducted investigation and preparation of the original draft under the supervision of Prof. Sergio (Yani) Carlos Garcia (S.C.G), Prof. Cameron Clark (C.C), Dr. Anna Chlingaryan (A.C) and Dr. Martin Correa-Luna (M.C). All authors devised conceptualisation and contributed to the review and editing process.

Chapter 3 has been published in *Smart Agriculture Technology* and is titled 'A data-driven approach for optimising supplement allocation to individual lactating dairy cows in pasture-based systems'. All authors contributed to methods conceptualisation. Data curation, validation, and formal analysis was conducted by B.N.A under the supervision of A.C, C.C, M.C, and S.C.G. Data visualisation was conducted by B.N.A and A.C. The original draft was written by B.N.A and reviewed and edited by A.C, C.C, M.C, and S.C.G.

Chapter 4 has been published in *Artificial Intelligence in Agriculture* and is titled 'Leveraging artificial intelligence and evolutionary algorithms for optimising cow supplementation and milk production'. All authors contributed to methods conceptualisation. Data curation, formal analysis, and data visualisation was conducted by B.N.A. The original draft was written by B.N.A and reviewed and edited by A.C, C.C, M.C, and S.C.G.

Chapter 5 has been published in *Remote Sensing* and is titled 'Data augmentation and interpolation improves machine learning-based pasture biomass estimation from sentinel-2 imagery'. All authors contributed to methods conceptualisation. Data curation, validation, visualisation and formal analysis was conducted by B.N.A under the supervision of A.C, C.C, M.C, and S.C.G. The original draft was written by B.N.A and reviewed and edited by A.C, C.C, M.C, and S.C.G.

Chapter 6 has been published in *Smart Agriculture Technology* and is titled 'Machine learning for grazing event detection and pasture utilisation quantification from Sentinel-2 data'. All authors contributed to methods conceptualisation. Data curation, validation, visualisation and formal analysis was conducted by B.N.A under the supervision of A.C, C.C, M.C, and S.C.G. The original draft was written by B.N.A and reviewed and edited by A.C, C.C, M.C, and S.C.G.

Chapter 7 has been published in *Journal of Agricultural and Food Research* and is titled 'Transfer learning and stacking ensembles for biomass estimation from smartphone imagery in pasture-based dairy systems'. All authors contributed to methods conceptualisation. Data curation, formal analysis,

and data visualisation was conducted by B.N.A and A.C. The original draft was written by B.N.A and reviewed and edited by A.C, C.C, M.C, and S.C.G.

Chapter 8 of this thesis focuses on the general discussion, conclusions, and recommendations for future research.

References for each chapter are consolidated at the end of the thesis to help decrease the thesis size.

Blessing Nnenna Azubuike

December 30, 2025

Supervisor statement

I, as the supervisor for the candidature of Blessing Azubuike, confirm that the following authorship attribution statements provided in this thesis are accurate and reflect the contributions of all listed authors.

Prof. Sergio (Yani) Carlos García

December 30, 2025

GENERATIVE AI ATTRIBUTION STATEMENT

During the preparation of this thesis, I, Blessing Nnenna Azubuike, used Microsoft Copilot for the purposes of text enhancement and ChatGPT 4o (https://chat.openai.com), to assist in writing and debugging R-language (https://www.Rproject.org) and Python code (https://www.python.org) errors. The use of this generative AI tool includes spelling corrections, minor sentence restructuring, and clarity enhancement. The author confirms that where text was modified by generative AI, the content was reviewed for possible errors, inaccuracies, and bias. The author takes full responsibility for the submitted thesis, confirms the work is their own, and has used generative AI in accordance with University guidelines and policies .

DEDICATION

This thesis is dedicated to my parents who sacrificed everything for me to be where I am today and have never stopped being my greatest role models of all time.

Mr & Mrs Azubuike Udu Eku

ACKNOWLEGEMENT

I thank the Lord God Almighty, my rock, my greatest support system and my foundation. I thank him for making this dream possible. I could not have done this without Him. This is indeed the doing of the Lord. I am thankful for His unwavering love, mercy and grace granted me to pursue my academic career at the University of Sydney. Thank you for seeing me through this journey and making it a success.

I deeply appreciate my lead supervisor, Professor Sergio (Yani) Carlos Garcia, who has been not only a supervisor but a father figure here in Australia. Thank you for believing in me from that first email on 24 April 2020 when you agreed to be my supervisor, until this day as I write these words. Your mentorship has been invaluable in shaping both my academic path and personal growth. From the moment you welcomed me at Sydney airport, your generosity and steadfast support have meant everything. Thank you for always telling me I could do a PhD if I decided to try, and for the support and encouragement during COVID when I almost gave up on coming. I am especially thankful for the many opportunities you created throughout this journey, which broadened my perspective and deepened my sense of purpose. Thank you for your constructive feedback, critical insights, thoughtful discussions, steady encouragement, and for making everything look easy with a smile. If I decide to do another PhD, do not go anywhere, you will undoubtedly be my first choice once again.

I express my sincere appreciation to my auxiliary supervisor, Professor Cameron Clark, for his patience and guidance throughout my PhD. Your unwavering support was a true blessing, and I was fortunate to have you as my co-supervisor. I extend my sincere gratitude to Dr. Anna Chlingaryan for your invaluable guidance and encouragements, especially with my manuscripts and for rigorously validating my codes and analytical processes. I am profoundly grateful to Dr. Martin Correa-Luna for your steadfast support throughout this journey, especially for your insightful manuscript reviews and valuable constructive feedback.

To my ever-loving husband, my best friend and my number one fan, Udo: my deepest gratitude goes to you for being there for me, believing in me, and encouraging me every step of the way. I know you did not sign up for all of this, but thank you for embracing it and being the unwavering support, I could ever ask for. I remember you saying, "There are two awards that should be given at your convocation, one for you and one for me," and whilst I would laugh, I know you truly deserve it. I thank God for bringing you into my life. Here is to our unending love, and thank you once again.

To my beloved parents, Mum (Rare) and Dad (Diamond): words cannot fully express how much I owe you for everything you have invested in my career. I love you both deeply and am profoundly grateful for all you have done. Thank you for praying for me, caring for me, and believing in me throughout this journey. To my only brother, Samson Azubuike, thank you for your unwavering support and for always standing by me.

To Gold: it pains me to write this in your absence, but I know and am certain that you can see and read these words. I know you are smiling and happy. Thank you for supporting me through this journey and for interceding for me while you lived. You are not forgotten.

I extend my heartfelt gratitude to my entire extended family for standing by me throughout this journey and for your unwavering support through prayers and words of encouragement. I am profoundly grateful.

I express my gratitude to Mrs. Kirsty Jaensch for your kind support and availability throughout this journey, especially during times when Yani was unavailable. Thank you for always looking out for me.

I thank my special ladies at the Lodge B1 at Camden campus accommodation, especially Vivien Tan and Katerina Stasinopoulou, for all your support throughout this journey and most especially during the last few months of my PhD. Thank you for always taking care of me and my little baby. I cherish you all.

I thank my friends and colleagues at the Centre for Carbon, Water and Food (CCWF) especially Joseph, Damilola, Milad, Rezaul and Mulisa, for your brotherhood, unwavering support and constant encouragement throughout this journey. I extend my gratitude to the entire CCWF staff and Mrs Lynne Gardner for your invaluable help, support and availability.

I gratefully acknowledge the financial support provided by the DairyUP Scholarship for my tuition and overseas health insurance, the Thomas Lawrence Pawlett Scholarships for my living expenses, and the Faculty of Science Scholarship for supplementary research funding.

It has been an honour and a true blessing to pursue my PhD at the University of Sydney.

ABSTRACT

Efficient management of pasture-based dairy systems can benefit substantially from integrating multiple interdependent processes including individual cow supplementation, real-time pasture monitoring, and tactical grazing allocation, but traditional herd-level approaches cannot address this complexity with sufficient precision, failing to account for biological variation between individual cows, dynamic pasture growth, and climate variability. This thesis developed and validated various artificial intelligence and machine learning methods across three critical domains (precision feeding optimisation, pasture biomass estimation, grazing event detection), establishing conceptual and empirical foundations for integrated precision management in commercial dairy operations. The research commenced with a comprehensive literature review (Chapter 2) identifying five critical gaps comprising absence of empirical validation for precision feeding optimisation systems delivering milk yield gains, vegetation index saturation and cloud gaps limiting satellite remote sensing reliability, insufficient spatial and temporal validation, lack of cost-effective automated grazing event detection techniques, and absence of integrated decision support frameworks. Feeding optimisation was tackled through two studies. First, a Random Forest model trained on 130 Holstein-Friesian cows (32,504 records) was combined with Differential Evolution to reallocate concentrate to 81 cows across 91 days without increasing total daily allocation, achieving an 8% theoretical milk yield increase (Chapter 3). Second, four evolutionary algorithms (NSGA-II, SPEA-II, SMS-EMOA, RVEA) were applied to the more demanding problem of simultaneously maximising daily herd milk yield and minimising deviations in individual concentrate allocation, trained on 1,053 cows (456,078 records) and applied to 165 cows over 30 days (Chapter 4) . NSGA-II achieved the highest performance (8.64% more milk) with the fastest computation time, whilst SPEA-II achieved 7.94%, RVEA 7.41%, and SMS-EMOA 6.63%, with all four algorithms statistically validated across 10 independent runs. Collectively, these studies demonstrate that exploiting cow-to-cow biological variability through AI-driven supplementation can improve herd productivity without increasing the total feed budget, though controlled experimental validation under commercial feeding conditions remains necessary.

Accurate pasture biomass estimation is a prerequisite for the feeding optimisation framework developed in Chapters 3 and 4, as it quantifies the pasture available to each cow before supplementation decisions are made. The gap in reliable pasture biomass estimation was addressed through two complementary approaches, each targeting a distinct platform limitation. Ground truth for both approaches was provided by rising plate meter measurements collected weekly to fortnightly across paddocks on each farm. First, Sentinel-2 satellite imagery across 16 commercial farms (November 2021 to July 2024) was used to develop an XGBoost estimation model, where multiquadric interpolation resolved cloud-induced temporal gaps and progressive seasonal training maintained accuracy across changing conditions and seasons, achieving an $R^2 = 0.70$ and MAE = 216 kg DM/ha and outperforming the commercial Pasture.io platform (MAE = 240 kg DM/ha) on matched test observations (Chapter 5).

These satellite biomass estimates from Chapter 5 were then extended to automated grazing event detection across 12 farms (July 2022 to June 2024), with Random Forest achieving strong within-year detection performance (F1 = 0.878) and One-Class Support Vector Machine demonstrating superior cross-year transferability (F1 = 0.692), outperforming supervised models including Random Forest by 7.6% on Year 2 data despite those models experiencing an average 24.2% performance degradation (Chapter 6). Independent pasture utilisation validation on 218 GPS-confirmed events demonstrated strong biomass quantification accuracy (pre-grazing $R^2 = 0.966$, post-grazing $R^2 = 0.998$, removal $R^2 = 0.922$), establishing proof-of-concept that satellite systems can not only detect when grazing occurs but also accurately quantify biomass changes, though temporal resolution constraints prevented validation of 80.6% of events due to insufficient satellite observations bracketing grazing periods, requiring multi-sensor fusion for comprehensive monitoring. Second, ground-based estimation using smartphone images was investigated as a practical, low-cost and durable alternative to remote sensing, deployable across diverse farm environments without dependence on satellite availability or specialised equipment. Stacking ensemble methods combining frozen pre-trained convolutional neural network features with hand-crafted vegetation indices and weather variables were applied to 3,291 ground-level images from 15 farms, achieving more moderate accuracy ($R^2 = 0.561$, MAE = 351 kg DM/ha). Importantly, this approach maintained sensitivity across broader biomass ranges where satellite vegetation indices saturate, offering a complementary capability, though spatial generalisation to unseen farms remains limited (Chapter 7).

Collectively, this thesis demonstrates that methods developed across each of the three domains function independently when validated on commercial farms, establishing the three critical building blocks required for integrated systems. Yet the ultimate goal of a closed-loop system, where measured pasture consumption and grazing event records directly inform individual cow feeding decisions in real time, remains to be achieved, requiring advances in temporal resolution, spatial transferability, and system integration. This research provides the empirical foundations and identifies the specific technical requirements for developing unified decision support systems that translate algorithmic capability into demonstrated value for commercial dairy farming operations, supporting more efficient and productive dairy production systems. All code implementations are organised by chapter and openly available at https://github.com/Stansfash/thesis-python_codes.

Keywords: Artificial intelligence; Automated grazing event detection; Machine learning; Pasture biomass estimation; Pasture-based dairy systems; Precision feeding; Remote sensing; Transfer learning.

ABBREVIATIONS AND ACRONYMS

The following abbreviated terms have been used throughout the thesis and are defined at first used in each chapter. Abbreviations that are used exclusively in Tables or Figures are not listed here and are defined within their respective Tables or Figures.

AI.....Artificial Intelligence
ANN .....Artificial Neural Network
ANOVA.....Analysis of Variance
BCS .....Body Condition Score
BW.....Body Weight
CART .....Classification and Regression Trees
CCC .....Concordance Correlation Coefficient
CI.....Confidence Interval
CIVE.....Colour Index of Vegetation Extraction
CNN.....Convolutional Neural Network
CONC.....Concentrate Intake
CP .....Crude Protein
CSH .....Compressed Sward Height
CV .....Coefficient of Variation
DE.....Differential Evolution
DIM .....Days in Milk
DL.....Deep Learning
DM.....Dry Matter
DMI .....Dry Matter Intake
DNN .....Deep Neural Network
DenseNet121 .....Densely Connected Convolutional Networks
EA.....Evolutionary Algorithm
EVI .....Enhanced Vegetation Index
ExG.....Excess Green Index
FN.....False Negative
FP.....False Positive
GA .....Genetic Algorithm
GBM.....Gradient Boosting Machine
GE.....Grazing Event / Grazing Events
GLCM .....Grey Level Co-occurrence Matrix
GLI .....Green Leaf Index
GPS.....Global Positioning System

HSV ..... Hue Saturation Value

IQR ..... Interquartile Range

IoT ..... Internet of Things

JSON ..... JavaScript Object Notation

KNN ..... K-Nearest Neighbours

LBP ..... Local Binary Pattern

LOFO-CV ..... Leave-One-Farm-Out Cross-Validation

LSTM ..... Long Short-Term Memory

LW ..... Live Weight

LoRaWAN ..... Long Range Wide Area Network

MAE ..... Mean Absolute Error

MAPE ..... Mean Absolute Percentage Error

ME ..... Metabolisable Energy

ML ..... Machine Learning

MODIS ..... Moderate Resolution Imaging Spectroradiometer

MQ ..... Multiquadric

MSE ..... Mean Squared Error

MY ..... Milk Yield

NDF ..... Neutral Detergent Fibre

NDVI ..... Normalised Difference Vegetation Index

NGRDI ..... Normalised Green Red Difference Index

NIR ..... Near Infrared

NOL ..... Number of Lactations

NN ..... Neural Network

NSGA-II ..... Non-dominated Sorting Genetic Algorithm II

OCSVM ..... One-Class Support Vector Machine

P-BDS ..... Pasture-Based Dairy Systems

PB ..... Pasture Biomass

PLF ..... Precision Livestock Farming

PU ..... Pasture Utilisation

RBF ..... Radial Basis Function

RF ..... Random Forest

RGB ..... Red Green Blue

RGB-D ..... Red-Green-Blue-Depth

RGBVI ..... Red Green Blue Vegetation Index

RMSE ..... Root Mean Squared Error

RNN ..... Recurrent Neural Network

RPM ..... Rising Plate Meter

RS ..... Remote Sensing

RVEA ..... Reference Vector Guided Evolutionary Algorithm

ResNet50 ..... Residual Network 50-layer

SAR ..... Synthetic Aperture Radar

SBX ..... Simulated Binary Crossover

SCC ..... Somatic Cell Count

SD ..... Standard Deviation

SLSQP ..... Sequential Least Squares Programming

SMS-EMOA ..... S-Metric Selection Evolutionary Multi-Objective Algorith

SOL ..... Stage of Lactation

SPEA-II ..... Strength Pareto Evolutionary Algorithm II

SVM ..... Support Vector Machine

SWIR ..... Shortwave Infrared

SfM ..... Structure from Motion

TAE ..... Total Absolute Error

TL ..... Transfer Learning

TMR ..... Total Mixed Ration

TN ..... True Negative

TP ..... True Positive

UAV ..... Unmanned Aerial Vehicle

VARI ..... Visible Atmospherically Resistant Index

VGG16 ..... Visual Geometry Group 16-layer

ViT ..... Vision Transformer

XGBoost ..... Extreme Gradient Boosting

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

General Introduction

THE CHALLENGE

Every day, dairy farmers managing pasture-based systems make hundreds of tactical decisions determining which paddocks to graze, how much supplement to allocate individual cows, and when to adjust stocking rates, yet these decisions must account for unpredictable climate variability, heterogeneous pasture growth patterns, and substantial biological variation between cows (Moscovici Joubran et al., 2021; Rutten et al., 2013). These decisions carry considerable economic stakes, as pasture-based systems dominate milk production across temperate regions globally, deriving competitive advantage from low-cost grazed forage that minimises purchased feed costs (García et al., 2014; Roche et al., 2017). In Australia, where feed costs comprise 40 to 60% of total production expenses, the difference between efficient and inefficient pasture utilisation combined with targeted supplementation directly determines farm profitability (Clark et al., 2018; Purcell et al., 2016). Despite this need for precision, traditional management tools remain inadequate for managing this complexity, with flat-rate feeding strategies failing to account for individual cow requirements despite documented gains from targeted supplementation (André et al., 2010; Little et al., 2016), manual pasture biomass (PB) assessment remaining labour-intensive and temporally sparse despite pasture growth rates exceeding 50 to 100 kg DM/ha/day (Gard et al., 2024; Kallenbach et al., 2020), and quantifying actual pasture utilisation remaining impractical without labour-intensive monitoring or expensive sensor technologies (Chebli et al., 2022). These limitations prevent farmers from implementing precision management strategies that optimise feed allocation, maximise pasture utilisation, and improve herd productivity, creating opportunities for transformative approaches capable of addressing the multi-dimensional complexity inherent in pasture-based dairy production whilst remaining practically deployable in commercial operations.

FROM COMPLEXITY TO CAPABILITY THROUGH ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) and machine learning (ML) have revolutionised decision-making across industries including medicine, engineering, and finance by solving complex problems that exceed human cognitive capacity (Russell & Norvig, 2021). Within AI, ML algorithms enable predictive analytics by learning patterns from data, capturing non-linear relationships between variables for supervised classification and unsupervised clustering tasks (Sarker, 2021). Additionally, deep learning (DL), a specialised subset of ML, employs multi-layered convolutional neural networks to automate image recognition and feature extraction, replacing tasks previously requiring expert visual analysis (LeCun et al., 2015; Simonyan & Zisserman, 2015). Beyond prediction, evolutionary algorithms provide prescriptive optimisation by systematically exploring solution spaces to identify optimal resource allocation strategies, handling multi-objective problems where traditional linear programming requires predefined objective weights or priorities, limiting its capacity to simultaneously explore the

full range of trade-off solutions across competing objectives. (Notte et al., 2021; Usigbe et al., 2023). These capabilities are now transferring to agriculture, where ML is used to predict crop yields and disease outbreaks, deep learning automates weed detection and fruit quality grading, and evolutionary algorithms to optimise irrigation scheduling and fertiliser application, collectively reducing costs and manual labour whilst improving decision accuracy (Kamilaris & Prenafeta-Boldú, 2018; Liakos et al., 2018). In dairy systems specifically, these capabilities offer potential to address the identified management challenges faced in the industry, with machine learning enabling individualised prediction of cow responses for targeted supplementation, deep learning facilitating automated PB assessment from imagery, and evolutionary algorithms providing optimisation frameworks that account for individual cow biological variation (Brito et al., 2025; Cockburn, 2020; Das et al., 2023). The conceptual framework for integrated precision management (Fig. 1.1) illustrates the ideal system where PB estimates inform allocation decisions, automated detection quantifies pasture intake, and feeding recommendations adapt to pasture availability. Whilst the conceptual framework presents UAV platforms as a potential sensing modality given their high spatial resolution, this thesis focuses on satellite and smartphone-based approaches as more scalable and operationally accessible alternatives for commercial dairy systems. In this integrated framework, information flows continuously between three core components, remote sensing for pasture monitoring, ML for grazing event detection, and optimisation algorithms for feed allocation, creating a closed feedback loop where each element informs and refines the others. However, current applications operate independently without information exchange between these components, treating each domain as an isolated technology rather than as interdependent elements of a unified decision support system. Therefore, this thesis develops and validates artificial intelligence methods across these three interdependent domains (precision feeding optimisation, pasture biomass estimation, grazing event detection), establishing the individual empirical building blocks required for achieving such integrated systems.

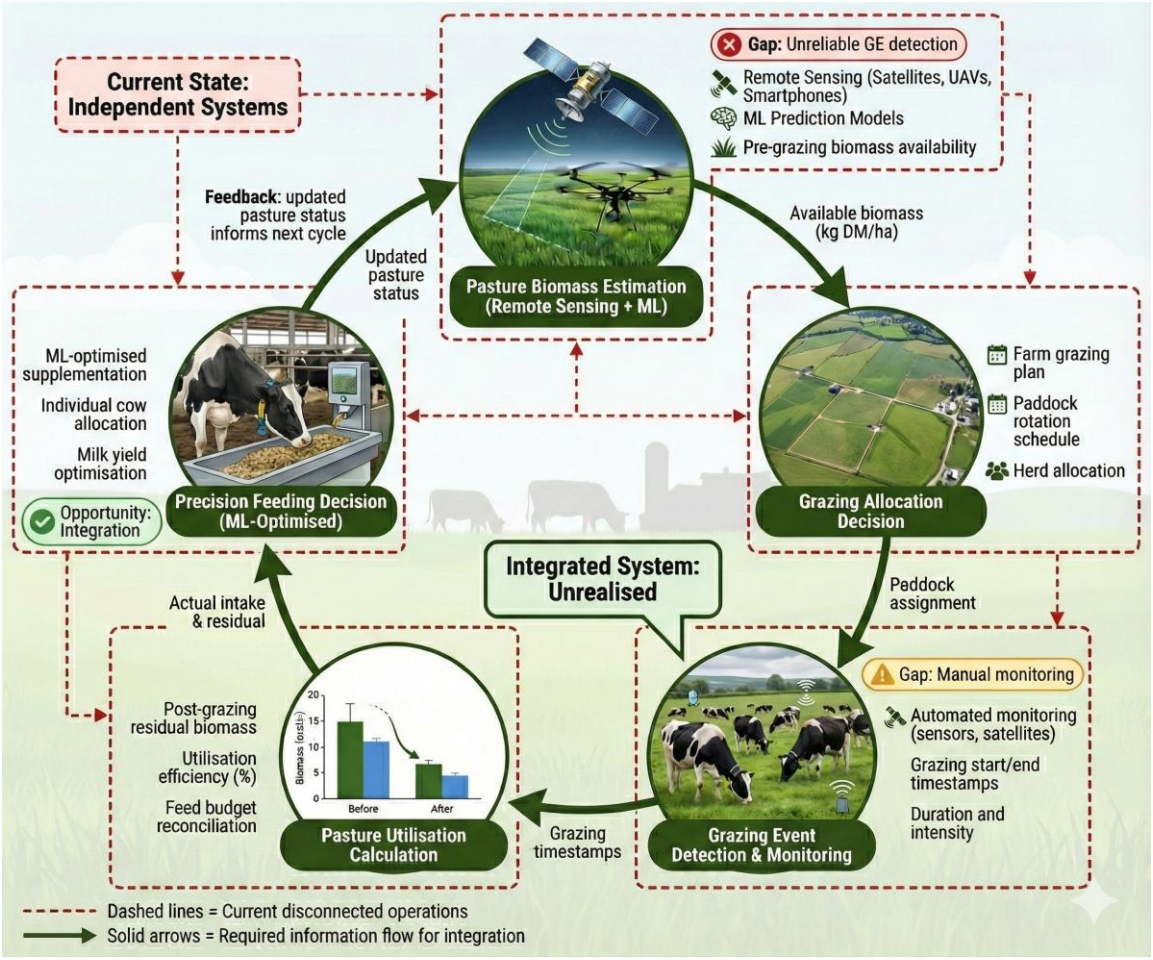


Fig. 1.1 Conceptual Framework for Integrated Precision Management.

Fig. 1.1 illustrates the conceptual framework where disjointed data streams, specifically PB estimates and animal nutritional requirements, are synthesised through ML algorithms. This architecture highlights the shift from reactive, labour-intensive observations to predictive, automated decision support, ensuring that grazing management and supplementary feeding are optimised simultaneously rather than in isolation.

The ideal closed-loop system, where outputs from each component continuously feedback to refine subsequent decisions, would integrate PB estimation using remote sensing and machine learning to provide pre-grazing availability, which would inform allocation determining rotation. Automated detection would monitor grazing recording timestamps. Utilisation calculation would quantify efficiency. Precision feeding would determine supplement allocation based on intake. Updated status would inform subsequent estimates. Currently these components operate independently. This thesis validates the individual components, establishing the empirical foundation required for achieving such integration.

RESEARCH GAPS

The literature review (Chapter 2) identifies critical knowledge gaps limiting practical implementation of artificial intelligence in pasture-based dairy systems. First, individualised feeding responses have been established through biological experimentation (García et al., 2007), demonstrating that cows respond differently to supplementation based on their physiological state and genetic merit. However, whilst machine learning and evolutionary algorithms show theoretical potential for precision feeding optimisation, these computational methods have not been empirically validated to quantify actual production gains and feed cost savings when applied to individual cow supplementation decisions under commercial conditions. Second, questions remain regarding the transferability of machine learning models for pasture PB estimation across diverse locations, environments, farms, and paddock scales, with most validation confined to specific sites raising concerns about spatial generalisability. Third, satellite-based estimation faces persistent challenges including vegetation index saturation at high PB levels and cloud interference limiting temporal coverage during critical periods. Fourth, quantifying actual pasture utilisation by grazing animals remains impractical for intensive rotational systems where short residence times (1 to 3 days per paddock) and frequent cloud cover prevent reliable automated detection of grazing events. These gaps motivated the thesis objectives to develop and validate artificial intelligence methods addressing each limitation, and establishing empirical foundations for their application in commercial dairy systems.

AIM AND OBJECTIVES

This thesis aims to develop and validate artificial intelligence methods across three interdependent domains (precision feeding optimisation, pasture biomass estimation, automated grazing detection), establishing empirical foundations for their application in commercial pasture-based dairy systems.

Objectives:

1. Develop and validate machine learning combined with evolutionary algorithms for precision feeding optimisation, quantifying production gains and feed cost savings compared to flat-rate supplementation under commercial conditions (Chapters 3 & 4).

2. Develop and evaluate PB estimation approaches addressing current limitations: (a) satellite-based methods with temporal interpolation to overcome cloud interference and vegetation index saturation, and (b) smartphone-based methods with transfer learning to maintain sensitivity at high PB, validating transferability across diverse farms and environments (Chapters 5, 7).

3. Develop automated frameworks for detecting grazing events and quantifying pasture utilisation from satellite data and interpolated biomass estimates developed in Chapter 5, evaluating prediction accuracy and spatial transferability to establish feasibility for commercial application (Chapter 6).

4. Synthesise findings across the three domains, discuss pathways towards integrated precision management systems, identify limitations, and propose future research directions for enhancing decision-making, production efficiency and profitability in pasture-based dairy systems (Chapter 8).

THESIS STRUCTURE

Chapter 2: Reviews artificial intelligence applications in pasture-based dairy systems and identifies research gaps.

Chapter 3: Integrates machine learning with evolutionary algorithms for precision feeding optimisation.

Chapter 4: Extends Chapter 3 by comparing multi-objective evolutionary algorithms for feeding optimisation at an individual cow level.

Chapter 5: Utilises temporal interpolation techniques to address cloud interference and vegetation index saturation, enhancing predictive accuracy of satellite-based PB estimation.

Chapter 6: Builds on the interpolated satellite biomass data from Chapter 5 to develop and evaluate machine learning frameworks for automated grazing event detection and pasture utilisation quantification across commercial farms.

Chapter 7: Investigates smartphone-based ground-level imagery as a practical, low-cost complement to satellite-based estimation, applying transfer learning and stacking ensemble methods for pasture biomass prediction and to address spatial transferability across diverse farm environments.

Chapter 8: Synthesises key findings, discusses pathways towards integrated precision management systems, and proposes future research directions.

CHAPTER 2

Literature Review

Pasture-based dairy systems are driven by the relentless challenge of high feed costs often comprising up to 60 percent of operational expenses. Profitability hinges on replacing traditional herd level management with precision strategies that account for individual cow variability and dynamic pasture availability. This literature review synthesises current artificial intelligence and machine learning applications for individualised feeding, remote sensing, pasture biomass estimation and grazing management, establishing the critical research gaps that Chapters 3 to 7 are designed to address.

ABSTRACT

Pasture-based dairy systems face management complexity arising from climate variability, dynamic pasture growth patterns, and individual cow behaviour. Artificial intelligence (AI) and machine learning (ML) applications are increasingly common to address this complexity. This review synthesises AI and ML applications across precision feeding optimisation, pasture biomass estimation through remote sensing, and grazing management, identifying critical research gaps constraining deployment of precision management systems. Unlike domain-specific reviews, this work examines how unresolved limitations within each domain collectively constrain progress toward integrated precision management systems in P-BDS. Traditional flat-rate feeding ignores individual cow variability despite documented gains from individualised supplementation, whilst linear regression models cannot capture non-linear dose-response relationships. Advanced ML algorithms demonstrate superior predictive performance, but integration with evolutionary algorithms lacks validation that automated systems quantify actual milk yield increases by leveraging cow-to-cow variability under commercial conditions. For pasture biomass estimation, satellite remote sensing combined with ensemble ML achieves accuracy, yet vegetation index saturation in dense pastures and cloud-induced temporal gaps constrain reliability. Ground-level computer vision approaches can address these limitations through deep learning architectures and pre-trained convolutional neural networks, maintaining sensitivity across broader biomass ranges despite limited training data. Monitoring pasture utilisation depends on expensive wearable sensors or labour-intensive manual observations, whilst satellite-based automated detection faces temporal resolution limitations. This review identifies that few studies integrate these three domains into unified decision support systems where real-time biomass estimates inform grazing allocation decisions triggering individualised supplementation. Future research should validate automated feeding systems, develop integrated approaches combining ML and deep learning to overcome satellite limitations, establish cost-effective automated grazing event detection, and create unified platforms linking pasture availability with utilisation efficiency and cow-specific feeding requirements.

Keywords: Artificial intelligence; Grazing management; Machine learning; Pasture biomass estimation; Precision feeding; Remote sensing.

INTRODUCTION

The global dairy industry is facing challenges in meeting rising demand whilst addressing climate change, resource constraints, and sustainability imperatives (Britt et al., 2018; Monteiro et al., 2024). Traditional herd-level management cannot adequately address complex interactions between individual animal variability, dynamic pasture growth, and fluctuating environmental conditions (Brito et al., 2025; Distante et al., 2025; Lokhorst et al., 2019; Melak et al., 2024; Rutten et al., 2013). Integration of artificial intelligence and machine learning within precision livestock farming frameworks enables transition from herd-level generalisations toward continuous, individualised, real-time decision support, reshaping responses to variability in animal performance, feed availability, and environmental conditions (Kaur et al., 2023; Kleen & Guatteo, 2023). Artificial intelligence refers to computer systems capable of performing tasks typically requiring human intelligence including learning, reasoning, and decision-making (Russell & Norvig, 2021). Machine learning, a subset of AI, enables systems to learn patterns and relationships from data without being explicitly programmed (Janiesch et al., 2021; Sarker, 2021). Deep learning (DL) represents a further specialisation within ML, employing neural networks to extract complex representations from large datasets (Alzubaidi et al., 2023; LeCun et al., 2015).

Pasture-based dairy systems (P-BDS), predominant in New Zealand, Australia, Ireland, and regions of Europe and South America, present unique AI implementation challenges compared to housed or confined systems (Correa-Luna et al., 2024; Moscovici Joubran et al., 2021; Murphy et al., 2022). Unlike confined systems where feed availability remains relatively controllable, P-BDS contend with unpredictable climate variability, heterogeneous soil characteristics, diverse pasture composition, and individual animal behaviour that interact in complex ways (Gargiulo et al., 2020; Shine & Murphy, 2022). Optimal decisions depend simultaneously on current pasture availability, anticipated pasture growth trajectories, grazing history, and cow-specific nutritional requirements. This complexity exceeds traditional decision-making capabilities, positioning P-BDS as systems where AI-driven solutions could deliver substantial improvements in productivity and profitability (Cabrera, 2025; Cabrera & Fadul-Pacheco, 2021; De Rosa et al., 2021; De Vries et al., 2023; García et al., 2020; Mahato & Neethirajan, 2024; Norton & Berckmans, 2017). AI-enabled precision management has evolved rapidly since 2015, driven by advancing sensor capabilities, wireless infrastructure, computational power, and proliferation of satellite platforms for pasture monitoring (Correa-Luna et al., 2024; Kamilaris & Prenafeta-Boldú, 2018; Morrone et al., 2022). ML applications in dairy farming have evolved substantially over more than two decades, progressing from simple rule-based decision support toward sophisticated data-driven systems integrating diverse sensor streams and optimisation algorithms (Shine & Murphy, 2022). These technologies enable continuous automated monitoring through wearable accelerometers, cameras, microphones, and remote sensing techniques (Aquilani et al., 2022; Kleen & Guatteo, 2023; Norton & Berckmans, 2017). The integration of remote sensing with

ML algorithms enables real-time optimisation of individual animal performance whilst maintaining herd productivity and sustainability (Kaur et al., 2023; Tzanidakis et al., 2023).

Three interdependent domains form precision management foundations in P-BDS, namely precision feeding optimisation, pasture biomass (PB) estimation, and grazing management. These domains interconnect through precision management loops where accurate PB estimation informs grazing allocation, automated grazing event detection enables calculated pasture utilisation efficiency, and integration of pasture availability with utilisation patterns triggers individualised supplementation to optimise milk yield (MY) whilst minimising feed costs. Currently, these domains operate independently in research and practice, preventing realisation of integrated systems that could transform operational efficiency in commercial P-BDS, thus, the rationale or focussing on these domains.

First, grain-based supplementary concentrate comprises 40-60% of production costs, but approximately 78% of Australian and New Zealand farms prefer flat-rate allocation strategies (equal distribution of feed for all cows) because of the ease of implementation, yet this method fails to account for individual cow requirements, leading to inefficient resource utilisation and suboptimal milk production outcomes (Akintan et al., 2025; Cabrera & Fadul-Pacheco, 2021; Clark et al., 2018; Craig et al., 2022). Individualised supplementation strategy has been shown to increase milk solids yield by approximately 7% (García et al., 2007), yet implementation of this method remains constrained. Whilst this productivity gain represents a compelling economic incentive, its commercial viability depends on whether the additional infrastructure and operational costs associated with individualised feeding systems are offset by the milk revenue gains, a calculation that varies with farm scale, milk price, and concentrate cost, and one that requires careful consideration in pasture-based systems where supplements represent only a portion of total diet. The research gap lies in the reliance of conventional biological models on parametric regression approaches, whether linear or nonlinear, that require predefined functional forms and cannot automatically adapt to the complex, individual-specific relationships between supplementation levels and milk yield responses, nor adequately capture the intrinsic biological variability amongst individual cows across diverse farm conditions (Cockburn, 2020; Dela Rue & Eastwood, 2017; Raedts & Hills, 2024). Consequently, prediction errors from such models range from 1.62 to 3.19 kg DM d^{-1} across diverse farm conditions (Krizsan et al., 2014). Advanced ML algorithms, neural networks, and evolutionary algorithms can model these non-linearities, but current implementations rely on retrospective observational data rather than experimental manipulation of feed rates and very rare attempt of integration of ML and evolutionary algorithms to solve this problem. This prevents the quantification of actual MY gains achievable through ML-optimised feeding, necessitating controlled trials to establish causal dose-response relationships and validate predicted improvements (Campos et al., 2023; Das et al., 2023; Notte et al., 2021; Souza et al., 2022; Uyeh et al., 2019).

Second, accurate PB estimation remains fundamental to effective feed budgeting and grazing management decisions, yet conventional approaches including the use of rising plate meters and visual assessment suffer from labour intensity, limited coverage, and systematic bias (Morse-McNabb et al., 2023; Villalobos-Villalobos & WingChing-Jones, 2023). Technical barriers limit remote sensing adoption, particularly spectral saturation where vegetation indices plateau at approximately 3 tonnes DM ha^{-1}, and cloud cover disrupting temporal consistency (Dubovik et al., 2021; Mutanga et al., 2023). Current satellite ML models achieve R^2 of 0.60 to 0.87 and RMSE between 350 and 900 kg DM ha^{-1} across temperate systems, with cloud-induced gaps spanning multiple weeks in high-rainfall dairy regions. Satellite platforms including Sentinel-2 provide sub-weekly revisit at 10-20m resolution, Landsat missions contribute 16-day revisit at 30m resolution, whilst PlanetScope offers paddock-scale monitoring and MODIS provides daily coverage (Chen et al., 2024; Correa-Luna et al., 2024; Gargiulo et al., 2020; Morse-McNabb et al., 2023; Ogungbuyi et al., 2024; Radeloff et al., 2024). Unmanned aerial vehicle platforms equipped with multispectral and thermal sensors enable centimetre-resolution monitoring and ground-truthing of satellite observations for high-precision PB estimation through 3D photogrammetry and canopy height models, achieving accuracy levels comparable to traditional destructive sampling methods (Ji et al., 2023; Laplacette et al., 2025), but persistent limitations necessitate complementary methodologies. Challenges include ground-truth data scarcity, temporal mismatch between satellite revisit and rapid pasture growth, and spatial heterogeneity in mixed further limit prediction accuracy (Correa-Luna et al., 2024; Ogungbuyi et al., 2024). Computer vision approaches using smartphone imagery with ML, DL and transfer learning could potentially increase accuracy through flexible user-determined acquisition timing and frequencies and maintained sensitivity across broader PB ranges, and reduced cloud cover constraints (Kamilaris & Prenafeta-Boldú, 2018; McCarthy et al., 2022; Stumpe et al., 2023; Vahidi et al., 2023), but these methods would require standardisation and validation to establish whether ML approaches can deliver accuracy improvements sufficient for implementation in dairy farms.

Third, knowing available PB before grazing provides limited value without corresponding knowledge of actual consumption during grazing and residual PB remaining for regrowth. Monitoring grazing event and calculating pasture utilisation efficiency remain essential as the relationship between grazing intensity and subsequent pasture productivity exhibits strong non-linear characteristics, with both undergrazing and overgrazing compromising long-term pasture productivity through different mechanisms. Current approaches for monitoring grazing event and calculating utilisation patterns rely predominantly on labour-intensive manual observations or expensive wearable sensor technologies despite achieving high accuracy cost several hundred dollars per collar constrain adoption (Chebli et al., 2022; Lamanna, Bovo, & Cavallini, 2025; Nyamuryekung'e et al., 2023). Meanwhile, satellite-based approaches offer potential for low-cost, non-invasive monitoring through detection of vegetation change patterns, but insufficient temporal resolution for short-duration rotational grazing cycles (1-3

days) prevents reliable detection of individual grazing events, with cloud-induced gaps spanning multiple weeks in major dairy farms. Whilst manual recording of grazing events at the time of paddock allocation could theoretically address this challenge, in practice such records are inconsistently maintained on commercial farms, and distinguishing grazing-induced changes from natural senescence, drought stress, or meteorological effects in satellite time series requires sophisticated ML classification techniques (Holtgrave et al., 2023). Cost-effective automated systems combining satellite vegetation indices with ancillary data could enable calculation of pre-grazing and post-grazing PB levels for quantifying utilisation efficiency.

Common challenges pervade these three domains including data scarcity for training robust predictive models, inappropriate validation methodologies producing optimistic performance estimates that fail to generalise, integration complexity across heterogeneous data streams and sensor modalities, and economic barriers constraining commercial adoption particularly for small to medium-scale operations (De Vries et al., 2023). Critically, limitations including non-linear feeding response prediction (Dela Rue & Eastwood, 2017), spectral saturation and temporal gaps in satellite data (Morse-McNabb et al., 2023), and cost-effective techniques prevent development of integrated precision management systems. Resolving these challenges is essential to build unified decision support systems frameworks where accurate real-time PB estimation can inform tactical paddock allocation, automated grazing detection can quantify actual pasture utilisation efficiency, and integration pasture availability with utilisation patterns can trigger cow-specific supplementation optimising MY production whilst improving feed use efficiency and contributing to farm profitability.

The aim of this review is to synthesise, analyse and integrate current knowledge on AI and ML applications and capabilites in precision feeding optimisation, PB estimation, and grazing management within P-BDS, to examine the limitations that hinder model performance and integrations, and to identify critical research priorities and future directions. Key requirements include AI-driven feeding strategies, overcoming satellite constraints to improve predictive accuracy, and developing integrated, cost-effective decision support systems frameworks. Addressing these challenges would provide a clear pathway for transitioning from experimental prototypes to integrated precision management platforms delivering measurable improvements in production efficiency and sustainability for P-BDS globally.

LITERATURE IDENTIFICATION AND SELECTION

This review was conducted as a narrative review synthesising current knowledge on AI and ML applications across three interdependent domains in P-BDS: precision feeding optimisation, pasture biomass estimation through remote sensing, and grazing management. Literature was identified through searches of Scopus, Web of Science, Google Scholar, Semantic Scholar, and Elicit, supplemented by

direct searches of relevant journal and publisher websites. Searches were conducted using combinations of terms spanning the three core domains as summarised in Table 2.1. Publications from 2015 onwards were prioritised to reflect the period of accelerated ML and satellite platform development, with pre-2015 references retained where they represented foundational contributions or the only available evidence on a specific topic.

Peer-reviewed original research articles and review papers published in English were eligible for inclusion. Studies were included where they presented empirical data from experiments, field measurements, or model testing, proposed or evaluated a new method, algorithm, or framework relevant to the three core domains, or synthesised evidence directly applicable to P-BDS. Non-dairy and non-pasture studies were included where the methodology or technology was directly transferable to P-BDS contexts. Studies were excluded where they lacked peer review, were published in languages other than English, or had no direct relevance to the three core domains. Final inclusion was determined through full-text reading and assessment of relevance to the review objectives.

Table 2.1 Literature search terms used across databases.

Domain	Search Terms
General	Artificial intelligence; machine learning; deep learning; precision livestock farming
Precision Feeding	Precision feeding; individualised feeding; individualised supplementation; flat-rate feeding; feed optimisation; evolutionary algorithms; genetic algorithms; feed intake prediction; dry matter intake; ration formulation; linear programming dairy; concentrate supplementation; milk yield prediction; surrogate model dairy; mathematical optimisation models in dairy
Pasture Biomass Estimation	Pasture biomass estimation; pasture monitoring; remote sensing; satellite imagery for biomass estimation; vegetation indices; NDVI; spectral saturation; UAV; computer vision biomass estimation; deep learning for pasture biomass estimation; convolutional neural network biomass; image classification; smartphone pasture monitoring; compressed sward height; dry matter estimation; biomass prediction; Sentinel-2; challenges in remote sensing for pasture biomass estimation; rising plate meter
Grazing Management	Grazing management; pasture utilisation; pasture utilisation efficiency; grazing event detection; rotational grazing; GPS collar; wearable sensor livestock; satellite grazing detection; automated grazing monitoring; machine learning grazing detection; SAR grassland monitoring; optical satellite grazing
System Context	Pasture-based dairy systems; pasture-based dairying; dairy cattle; dairy cows; milk yield

DATA-DRIVEN APPROACHES IN PASTURE-BASED DAIRY SYSTEMS

Traditional approaches to modelling agricultural systems rely predominantly on mechanistic models derived from controlled experiments and theoretical understanding of biological processes. These model-driven approaches encode domain knowledge through mathematical equations describing relationships between inputs and output variables, enabling predictions based on first principles (Jones et al., 2017). However, mechanistic models face limitations when addressing complex systems characterised by non-linear interactions, spatial heterogeneity, and temporal variability inherent in P-BDS (Beukes et al., 2018; García et al., 2020). Data-driven approaches employ ML algorithms to automatically discover patterns and relationships directly from observational data without requiring explicit mathematical formulation of underlying processes (Liakos et al., 2018; Sharma et al., 2022). Rather than encoding expert knowledge through equations, these methods learn predictive models through exposure to examples, adapting to system complexity and capturing non-linearities that mechanistic approaches struggle to represent (Kamilaris & Prenafeta-Boldú, 2018). Fig. 2.1 contrasts fundamental characteristics distinguishing these paradigms. The shift toward data-driven methods in dairy systems has accelerated recently, driven by proliferation of sensor networks, satellite platforms, and computational resources enabling analysis of large-scale datasets (van Klompenburg et al., 2020).

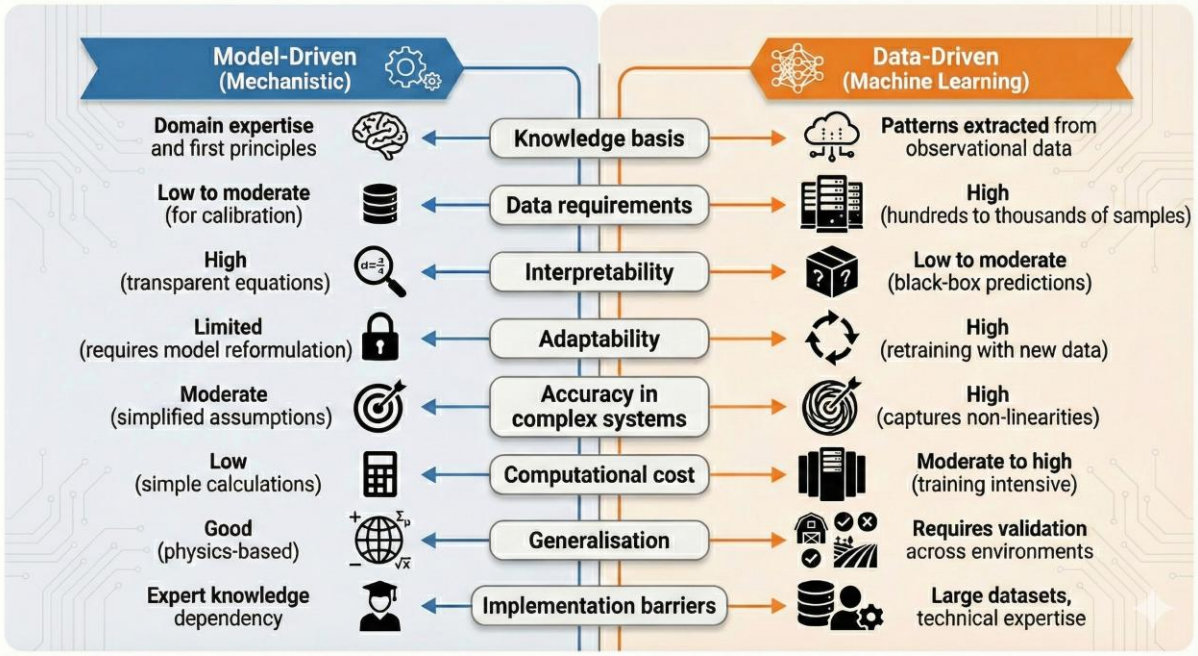


Fig. 2.1 A comparative analysis of Model-Driven (Mechanistic) versus Data-Driven (Machine Learning) approaches in dairy production systems.

A comparison of the strengths and limitations of mechanistic models (left) versus machine learning algorithms (right). The central column identifies eight criteria used to evaluate the suitability of each approach, highlighting the inverse relationship between data requirements and model interpretability.

Within this data-driven paradigm, techniques span several distinct categories. Supervised learning algorithms, including Random Forest, Gradient Boosting Machines, and Support Vector Machines, learn from labelled examples to predict continuous outcomes (regression) or discrete classes (classification), making them suited to tasks such as milk yield prediction and grazing event detection. Unsupervised learning methods, such as k-means clustering, identify structure in unlabelled data and are used for segmenting cow production groups or paddock management patterns. Nonlinear methods including tree-based ensembles and neural networks capture complex dose-response relationships that linear regression and logistic models cannot adequately represent. For resource allocation problems, combinatorial optimisation approaches include mathematical programming techniques such as linear and nonlinear programming, which guarantee optimal solutions within defined constraints. Alternatively, metaheuristic algorithms such as genetic algorithms and differential evolution use population-based search to find high-quality solutions in large, non-convex search spaces where exact methods become computationally intractable.

Published implementations of ML in dairy production demonstrate common workflow patterns documented across recent reviews (Liakos et al., 2018). Studies typically integrate data from remote sensing platforms, on-farm sensors, and manual ground truth measurements. Standard preprocessing addresses quality issues and normalisation before feature engineering extracts predictive variables such as vegetation indices and temporal patterns. Reported methodologies employ diverse ML algorithms and neural networks, with performance evaluated using cross-validation and independent test datasets (Roberts et al., 2017). However, robust model performance in agricultural datasets requires validation strategies that account for spatial and temporal data structure, as standard random cross-validation frequently produces optimistic estimates that fail to generalise across farms and seasons. Practical implementations described in literature integrate trained models with decision support platforms, though long-term monitoring and retraining protocols remain underreported in published studies. Despite demonstrated capabilities, adoption barriers constrain implementation of data-driven approaches in commercial dairy systems. Primary challenges include scarcity of labelled training data particularly for specialised applications, computational infrastructure requirements exceeding resources available to small-scale operations, technical expertise gaps limiting farmer interpretation and trust in algorithmic recommendations (Hills et al., 2015). Validation concerns also arise where inappropriate methodologies produce optimistic performance estimates failing to generalise across diverse farm conditions and scenarios. Economic considerations including initial investment costs, uncertain return on investment timelines, and maintenance requirements further limit adoption particularly amongst resource-constrained producers (Pivoto et al., 2018). Researchers have proposed transfer learning approaches to reduce data requirements, cloud-based platforms to lower infrastructure barriers, and improved validation frameworks, though widespread implementation of these solutions remains limited in commercial practice (van Klompenburg et al., 2020).

REMOTE SENSING AND ARTIFICIAL INTELLIGENCE FOR PASTURE BIOMASS ESTIMATION

Knowing how much pasture is available in each paddock at any given time is essential to effective tactical decision-making in P-BDS, influencing grazing rotation schedules, stocking rate adjustments and supplementary feeding requirements. However, the accuracy and consistency of this knowledge varies substantially depending on the measurement approach employed. Traditional approaches relying on rising plate meter and destructive sampling provide point estimates of standing PB through labour-intensive field measurements that constrain spatial coverage, introduce observer variability and impose temporal lags between assessment and management action (Morse-McNabb et al., 2023; Villalobos-Villalobos & WingChing-Jones, 2023). The proliferation of satellite platforms offering frequent multispectral imagery at paddock-relevant spatial resolutions, combined with advances in AI algorithms capable of extracting quantitative PB estimates from complex spectral signatures, has catalysed a paradigm shift towards continuous, automated pasture monitoring since 2015 (Ogungbuyi et al., 2024; Shahi et al., 2025). However, incorporating these technological capabilities into commercial dairy systems that consistently deliver actionable information under diverse environmental conditions and management scenarios remains an incompletely solved challenge. Despite a decade of research demonstrating technical feasibility in controlled trials, commercial adoption remains minimal, with the vast majority of dairy farmers worldwide continuing to rely on visual assessment or periodic manual measurements rather than automated continuous monitoring systems (De Vries et al., 2023). This gap between experimental demonstration and practical implementation reflects fundamental limitations in current research methodologies and deployment strategies that prevent translation of algorithmic capability into demonstrated value for farm decision-making.

Satellite-Based Platforms and Multi-Sensor Integrations

Satellite-based remote sensing platforms now provide opportunities for continuous, non-invasive pasture monitoring across multiple spatial and temporal resolutions. For example, the Sentinel-2 constellation delivers multispectral imagery at 10 to 20 m spatial resolution with sub-weekly revisit times, enabling paddock-scale PB estimation that was previously unattainable with coarser resolution sensors such as MODIS or early Landsat missions (Cândido et al., 2025; Chen et al., 2021; Radeloff et al., 2024). The 13 spectral bands currently offered by Sentinel spanning visible to shortwave infrared wavelengths provide rich spectral information for vegetation characterisation, with particular value derived from red-edge bands at 705, 740 and 783 nm that demonstrate enhanced sensitivity to chlorophyll content and canopy structure compared to conventional red and near-infrared bands utilised in traditional Normalised Difference Vegetation Index (NDVI) calculations (Askari et al., 2019; Ogungbuyi et al., 2024). Complementary platforms including Planet CubeSat nanosatellites offer daily

imagery at 3 to 5 m resolution, enabling detection of rapid PB changes between Sentinel-2 and addressing temporal continuity challenges inherent in single-platform approaches, though at substantially higher data acquisition costs through commercially available platforms (Correa-Luna et al., 2024; Gargiulo et al., 2020; Morse-McNabb et al., 2023; Ogungbuyi et al., 2024). Gargiulo et al. (2020) demonstrated that the integration of electronic rising plate meter measurements with Planet CubeSat and Sentinel-2 satellite data across commercial dairy farms in New South Wales enabled spatial and temporal PB estimation, with daily Planet imagery detecting rapid PB changes between Sentinel-2 acquisition dates. The multi-platform integration substantially improved estimation accuracy compared to single-sensor approaches, particularly during periods of rapid pasture growth or cloud interference affecting individual platforms.

The integration of proximal sensing measurements with satellite observations substantially enhances model calibration and validation whilst enabling multi-scale PB monitoring strategies. Cândido et al. (2025) demonstrated that the fusion of ultrasonic sensor measurements from the PaddockTrac proximal sensing system with vegetation indices derived from Landsat 7 and Sentinel-2 imagery achieved superior performance compared to satellite-only approaches, with XGBoost models attaining (Coefficient of determination; $R^2 = 0.86$ and Root mean square error; RMSE = 538 kg ha^{-1}) across multiple growing seasons at a single experimental site under homogenous pasture conditions in Missouri, USA. The complementary nature of ground-based and satellite measurements addresses fundamental scaling challenges in pasture monitoring, with proximal sensors providing high-frequency spot measurements for model calibration. In turn, satellite platforms deliver spatially comprehensive coverage enabling whole-farm PB mapping and identification of within-paddock heterogeneity critical for precision grazing management. Spatial mapping of PB distribution within paddocks enables identification of within-paddock heterogeneity, with the potential to support targeted grazing allocation decisions that prevent overgrazing of sparse areas whilst ensuring efficient utilisation of high-biomass zones, thereby improving pasture utilisation efficiency and sward quality (Correa-Luna et al., 2024; Freitas et al., 2022; Gargiulo et al., 2023).

Machine Learning Algorithms for Spectral-Biomass Estimation

Machine learning algorithms have emerged as the dominant analytical framework for transforming multispectral satellite data into quantitative PB estimates, superseding traditional linear regression approaches that are unable to capture complex, non-linear relationships between spectral reflectance and vegetation biophysical properties (De Rosa et al., 2021; Stumpe et al., 2023; Zhou et al., 2025). Among these ML approaches, tree-based ensemble methods and gradient boosting machines demonstrate particular robustness across diverse environmental conditions and management systems, leveraging ensemble learning principles to reduce overfitting whilst maintaining high predictive

performance (Cândido et al., 2025; Morse-McNabb et al., 2023). Recent studies have reported R^2 values ranging from 0.60 to 0.87 with RMSE between 350 and 900 kg DM ha^{-1}, with the lower end of this error range considered sufficient for tactical management applications whilst higher error levels constrain the precision required for individual paddock allocation decisions. These figures however reflect studies conducted across different climate zones, pasture species, and validation approaches, ranging from temperate Australian dairy pastures evaluated with neural networks (Chen et al., 2021) to tropical Brazilian pastures assessed using support vector regression and random forest algorithms (Fernandes et al., 2024). These ensemble approaches capitalise on the heterogeneity of pasture systems by combining multiple weak learners into robust predictive frameworks capable of handling non-linear spectral responses, seasonal variations and management-induced heterogeneity that constrain simpler regression models. The algorithms learn hierarchical feature representations from multispectral data, automatically identifying relevant spectral combinations and vegetation indices without requiring explicit specification of functional forms. This data-driven adaptability proves particularly valuable in mixed-species pasture systems where species composition, phenological stage and environmental stress patterns introduce spectral complexity that defies parametric modelling approaches. Different grass species exhibit distinct chlorophyll concentrations, leaf structures and canopy architectures that generate unique spectral signatures. Meanwhile, temporal variation in growth stages creates shifting reflectance patterns within individual paddocks that cannot be adequately captured by fixed linear relationships between spectral bands and biomass (Prasad et al., 2023). However, the black-box nature of these ensemble methods constrains interpretability and limits capacity to diagnose model failures or incorporate domain knowledge regarding plant physiology and pasture ecology into prediction frameworks (Welchowski et al., 2022).

Spectral Saturation and Temporal Discontinuity Constraints

Despite substantial methodological advances with remote sensing, satellite-based PB estimation faces persistent barriers that constrain operational deployment in P-BDS. The challenge of spectral saturation in vegetation indices at high biomass levels remains incompletely resolved, reducing model sensitivity precisely within the high biomass range where tactical grazing decisions are most critical (Askar et al., 2018; Ogungbuyi et al., 2025). NDVI particularly show pronounced saturation at high leaf area index values, resulting from the convergence of near-infrared reflectance towards maximum values whilst red reflectance approaches minimum levels in dense canopy conditions (Carlson & Ripley, 1997; Mutanga et al., 2023). This saturation prevents reliable estimation in high-productivity systems where biomass frequently exceeds threshold values during optimal growing conditions. When NDVI saturates at approximately 0.8-0.9, further biomass increases produce negligible changes in the vegetation index value, rendering the sensor insensitive to the precise biomass levels required for tactical grazing

management decisions in high-productivity pastures where optimal harvest timing depends on discriminating between 2,500 and 3,500 kg DM ha^{-1}(Dubovik et al., 2021).

Alternative spectral formulations offer partial mitigation of saturation effects through incorporation of red-edge and shortwave infrared wavelengths that maintain sensitivity across broader biomass ranges. Askar et al. (2018) demonstrated that the Normalised Difference Index 45, computed from Sentinel-2 band 5 (red edge at 705 nm) and band 4 (red at 665 nm), substantially outperformed traditional NDVI for forest biomass estimation, achieving R^2 of 0.79 compared to 0.44 for conventional formulations. Similarly, integration of shortwave infrared bands sensitive to vegetation water content and canopy structure provided enhanced discrimination of high-biomass conditions where optical vegetation indices saturate. SWIR wavelengths respond to three-dimensional canopy architecture and plant water status rather than leaf chlorophyll content, maintaining sensitivity as biomass increases because canopy structural complexity and total water content continue to change even when leaf area index reaches maximum values (Peng et al., 2023; Vahidi et al., 2023). However, these improvements remain incremental rather than transformative, with comprehensive solutions requiring integration of multiple spectral features through sophisticated ML architectures capable of learning complex, multidimensional spectral-biomass relationships beyond capabilities of simple vegetation index formulations (Mike et al., 2024; Shahi et al., 2025).

Data continuity gaps arising from cloud cover, atmospheric interference and fixed satellite revisit times equally hinder effective pasture monitoring systems. Temperate coastal climates and high-rainfall regions characteristic of major dairy production zones experience persistent cloud cover that can result in prolonged gaps spanning multiple weeks between usable satellite observations. This disrupts the temporal continuity required for near real-time decision support applications (Chen et al., 2021; Dubovik et al., 2021; Ogungbuyi et al., 2024). The temporal resolution of Sentinel-2 (five-day revisit at equator, longer at higher latitudes) proves insufficient to capture rapid pasture growth dynamics during optimal conditions where biomass accumulation can exceed 50 to 100 kg DM ha^{-1} per day in temperate climates. This necessitates temporal interpolation methods or integration of Synthetic Aperture Radar (SAR) data capable of penetrating cloud cover (Holtgrave et al., 2023; Vahidi et al., 2023). Vahidi et al. (2023) demonstrated that fusion of Sentinel-1 SAR with Sentinel-2 optical imagery provided improved temporal coverage and PB estimation accuracy compared to optical-only approaches, though challenges remain regarding SAR sensitivity to soil moisture and surface roughness that can confound vegetation signal extraction in intensively managed pasture-based systems.

UAV platforms equipped with multispectral, hyperspectral or thermal sensors enable centimetre to metre-scale spatial resolution monitoring. This bridges the gap between ground-based measurements and satellite observations, providing enhanced capability for detecting within-paddock heterogeneity

and validating coarser resolution satellite-derived estimates (Cho et al., 2007; Herrmann et al., 2011; Ji et al., 2023; Laplacette et al., 2025; Togeirode Alckmin et al., 2022). These systems offer operational flexibility in acquisition timing unconstrained by satellite orbital schedules, enabling targeted data collection under optimal atmospheric conditions or at critical phenological stages where rapid PB changes demand high temporal frequency monitoring (Askari et al., 2019; Lussem et al., 2020; Pullanagari et al., 2015; Vahidi et al., 2023). Three-dimensional photogrammetry and canopy height model generation from UAV imagery provide structural information complementary to spectral features, achieving biomass estimation accuracy levels comparable to destructive sampling methods through integration of vegetation volume with spectral indices (Bendig et al., 2015; Castro et al., 2020; Punalekar et al., 2018; Thomson et al., 2023). UAV-mounted RGB cameras, whilst considerably more affordable and simpler to operate, are constrained to visible wavelengths and lack the near-infrared and red-edge sensitivity required for reliable biomass prediction above approximately 3,000 kg DM/ha. As a result, multispectral or hyperspectral sensors remain preferable for quantitative biomass estimation despite their substantially higher cost and processing demands. Integration of hyperspectral reflectance data with ML algorithms enables estimation not only of biomass quantity but also nutritive value parameters including crude protein content, digestibility, and fibre fractions that critically influence livestock production outcomes but remain largely invisible to conventional multispectral approaches (Capolupo et al., 2015).

However, several constraints limit widespread adoption of UAV technologies for routine pasture monitoring in commercial P-BDS. Flight regulations, weather dependencies and limited spatial coverage constrain scalability of UAV operations to whole-farm monitoring applications, with typical mission endurance restricting coverage to tens rather than hundreds of hectares without multiple battery changes or deployment of expensive fixed-wing platforms (Bazzo et al., 2023; Gargiulo et al., 2023). Additionally, substantial capital costs for UAV platforms with research-grade multispectral or hyperspectral sensors, combined with operational expenses including insurance, maintenance, pilot training and data processing that substantially exceed ongoing costs of satellite data access. These combined costs create economic barriers particularly acute for small to medium-scale farming operations (Gargiulo et al., 2023; Stott et al., 2020).

Deep Learning and Computer Vision for PB Estimation

Ground-level computer vision technologies offer complementary capabilities to satellite remote sensing through provision of fine-scale PB estimation and pasture quality assessment at the paddock level. These approaches leverage the ubiquity of smartphones, digital cameras and vehicle-mounted imaging systems to enable frequent, low-cost monitoring that complements less frequent satellite data collection whilst providing assessment of pasture characteristics including species composition, flowering stage

and weed presence that is difficult to detect through multispectral satellite imagery alone (Freitas et al., 2022; Lussem et al., 2019; Woodrow et al., 2023).

Methodological approaches for ground-based PB estimation have evolved from traditional machine vision systems towards sophisticated DL architectures, reflecting broader technological advances in computer vision and the gradual accumulation of training datasets. McCarthy et al. (2022) developed a machine vision system using in-field cameras that detected pasture dry matter amount with r=0.715 and RMSE=381.5 kg DM ha^{-1} in predominantly perennial ryegrass paddocks through analysis of daily images compared with C-Dax measurements. C-Dax is a vehicle-mounted optical sensor system that uses active light reflectance to measure pasture mass rapidly across large areas, providing continuous spatial coverage along transects as the vehicle traverses paddocks (Ogungbuyi et al., 2024; Rennie et al., 2009). This traditional computer vision approach employed image analysis techniques to extract visual features correlated with biomass, demonstrating the feasibility of automated monitoring through fixed camera installations. Vahidi et al. (2023) advanced beyond conventional image analysis by developing artificial neural network models for PB estimation using ultra-high-resolution RGB imagery from UAVs, achieving robust performance across diverse paddock management conditions including rest, bale grazing and sacrifice areas through integration of canopy height models with spectral features. The shift from rule-based feature extraction towards data-driven learning of feature representations marked a methodological transition enabling more nuanced capture of complex relationships between visual patterns and PB characteristics.

However, the scarcity of comprehensive, labelled training datasets represents a fundamental constraint limiting development and validation of robust DL models for agricultural applications. Traditional supervised learning approaches requiring at a minimum scale thousands of annotated images prove economically and logistically prohibitive for P-BDS, where ground truth collection demands labour-intensive field measurements and destructive sampling protocols. This data scarcity bottleneck has driven adoption of transfer learning strategies as an enabling technology for agricultural computer vision, allowing development of high-performance models from limited domain-specific training data. Transfer learning enables leveraging hierarchical feature representations learned from large-scale image datasets such as ImageNet to initialise networks subsequently fine-tuned on limited agricultural imagery. This substantially reduces data requirements whilst maintaining prediction accuracy comparable to or exceeding conventional vegetation indices derived from multispectral imagery (Narayanan et al., 2021; U et al., 2022).

Woodrow et al. (2023) demonstrated the practical application of transfer learning through implementation of DenseNet121, a pretrained convolutional neural network architecture, for PB estimation from smartphone photographs across extensive grazing lands in Queensland. The

DenseNet121-based models achieved viable PB estimates, providing an accessible measurement approach for resource-constrained pastoral operations that cannot afford the labour costs or equipment investments required by traditional monitoring methods. The capacity to leverage pre-learned representations from millions of general images and adapt them to pasture-specific characteristics through fine-tuning on limited agricultural datasets represents a methodological breakthrough enabling deployment of sophisticated DL approaches without requiring prohibitive ground truth collection efforts. Feature extraction from intermediate convolutional layers of pretrained architectures including VGG16, ResNet50 and InceptionV3 captures hierarchical visual patterns spanning low-level texture and colour distributions through mid-level structural characteristics to high-level semantic representations. These encoded pasture attributes relevant to biomass estimation prove difficult to specify through hand-crafted features(Castro et al., 2020; Skovsen et al., 2019).

The integration of transfer learning with weak supervision strategies proves particularly valuable where annotated training data remains scarce and expensive to acquire through manual labelling processes. Narayanan et al. (2021) adapted VGG16 pretrained on ImageNet for multi-target regression of grass, clover and weed biomass percentages from paddock imagery, employing data augmentation and differential sample weighting to address missing labels and limited training set size. This approach achieved competitive performance using only 261 training images, demonstrating the viability of DL methodologies even for small agricultural datasets where traditional ML approaches might prove data-limited. The capacity to leverage partially labelled data through semi-supervised learning frameworks addresses a critical bottleneck in agricultural computer vision, where comprehensive ground truth collection across diverse species compositions, growth stages and environmental conditions would require prohibitive field effort. These systems enable rapid, non-destructive assessment that can be readily integrated into routine farm management workflows, offering potential for democratising precision pasture monitoring through ubiquitous consumer hardware rather than specialised sensing equipment.

Advanced DL architectures including 3D convolutional neural networks and attention mechanisms demonstrate emerging capabilities for temporal biomass prediction and quality assessment from sequential imagery, moving beyond single-timepoint estimation towards trajectory modelling and phenological stage classification. Castro et al. (2020) compared AlexNet, VGGNet and ResNet architectures for forage breeding programme applications, demonstrating that deep residual networks achieved superior performance for PB estimation from RGB orthoimages compared to conventional vegetation index approaches or shallower network architectures. The capacity of deeper networks to learn increasingly abstract feature hierarchies through residual connections that mitigate vanishing gradient problems enables extraction of subtle visual cues relating to pasture density, species composition and maturity that correlate with biomass accumulation and nutritive value characteristics

(He et al., 2016; Kamilaris & Prenafeta-Boldú, 2018). However, translating these architectural advances into farming systems faces persistent implementation constraints that constrain widespread adoption of these technologies.

Image acquisition consistency is a primary factor limiting operational deployment of computer vision approaches, with varying illumination conditions, camera angles, heights and timing relative to solar position introducing systematic variability. This confounds model predictions and reduces reliability under operational conditions (Lussem et al., 2019; McCarthy et al., 2022; Woodrow et al., 2023). Standardisation of capture protocols including camera height, angle and timing is essential for consistent feature extraction, though practical implementation across diverse farm environments and management workflows presents significant logistical challenges. Shadows, specular reflections and atmospheric conditions including haze or precipitation further degrade image quality and model performance, necessitating robust preprocessing and quality control procedures to identify and reject unsuitable acquisitions. The limited field of view inherent in ground-level imagery constrains spatial coverage compared to satellite or aerial platforms, requiring multiple images to be captured to characterise paddock-scale biomass distribution and introducing sampling considerations analogous to traditional field measurement approaches. Vehicle-mounted or smartphone-based imaging systems typically sample small areas (less than 1 m^2 per image) compared to paddock extents ranging from 1 to 20 ha in intensive dairy systems, necessitating strategic sampling designs to ensure representative coverage whilst maintaining practical data collection efficiency.

Despite these implementation challenges highlighted above, smartphone-based or digital cameras imaging systems offer compelling advantages including cost-effectiveness compared to specialised multispectral sensors, durability under field conditions, and ubiquitous availability enabling rapid deployment without substantial capital investment. Emerging smartphone capabilities including 3D photogrammetry through structure-from-motion algorithms and video capture enabling temporal sequence analysis present additional opportunities for enhanced PB estimation through integration of structural and dynamic information complementary to static RGB imagery. Three-dimensional reconstruction from multiple viewing angles captures canopy height and volume metrics analogous to LiDAR-derived measurements whilst video sequences enable detection of canopy movement patterns and temporal growth dynamics that static images cannot capture, potentially improving estimation accuracy through fusion of spatial structure with spectral characteristics.

The convergence of increasingly capable smartphone cameras, expanding availability of pretrained convolutional neural network architectures, and growing recognition of transfer learning as an enabling methodology for data-efficient model development positions ground-based computer vision as a complementary technology to satellite remote sensing for paddock-scale biomass monitoring. Critical

research priorities include establishment of standardised image acquisition protocols ensuring consistency across diverse environmental conditions, systematic evaluation of transfer learning approaches for adapting pretrained architectures to pasture-specific applications with minimal domain-specific training data, and exploration of multi-modal fusion frameworks integrating 2D imagery with 3D structural information and temporal video sequences. Addressing these challenges through coordinated research investment would enable smartphone-based computer vision systems to provide cost-effective, high-frequency biomass monitoring bridging temporal gaps inherent in satellite platforms whilst democratising access to precision pasture management technologies for resource-constrained farming operations. The diverse remote sensing platforms employed for PB estimation present distinct trade-offs amongst spatial resolution, temporal coverage, acquisition costs and operational constraints. Table 2.2 summarises the key characteristics, limitations and representative studies for satellite, UAV and ground-based monitoring approaches to facilitate platform selection based on specific operational requirements and resource constraints.

Table 2.2 Comparison of remote sensing platforms for pasture biomass estimation in dairy systems.

Sensors	Spatial Resolution	Temporal Resolution	Acquisition and Operational Cost	Coverage Area	Key Limitations	Papers
Sentinel-2	10-20 m	5 days (cloud dependent)	Free acquisition; operational costs limited to processing infrastructure	Continental (290 km swath)	Spectral saturation at high biomass, cloud cover dependency, coarse resolution for paddock-scale	(Chen et al., 2021; Gargiulo et al., 2020; Morse-McNabb et al., 2023; Ogungbuyi et al., 2024; Radeloff et al., 2024)
Landsat 8/9	30 m	16 days	Free acquisition; operational costs limited to processing infrastructure	Continental (185 km swath)	Lower temporal and spatial resolution, spectral saturation, cloud cover	(Ali et al., 2025; Cai et al., 2024; Chen et al., 2024; Nasiri et al., 2022)
PlanetScope	3-5 m	Daily	Commercial subscription; whole-farm monitoring potentially exceeding several thousand dollars annually	Global coverage	Radiometric calibration challenges, spectral saturation, cost barriers	(Cai et al., 2024; Dos Reis et al., 2020; Ogungbuyi et al., 2025)
UAV Multispectral	1-10 cm	User-defined (on-demand)	High equipment and operational costs including maintenance, insurance, pilot training and data processing	Farm-scale (hectares per flight)	Weather dependent, labour intensive, requires technical expertise, flight regulations, limited spatial coverage	(Alvarez-Mendoza et al., 2022; Bendig et al., 2015; Freitas et al., 2022; Ji et al., 2023; Laplacette et al., 2025; Togeirode Alckmin et al., 2022)
UAV RGB + SfM	1-5 cm	User-defined (on-demand)	Moderate equipment cost; lower than multispectral but similar operational demands	Farm-scale (hectares per flight)	Limited spectral information, processing complexity, weather dependent	(Lussem et al., 2019; Vahidi et al., 2023)

Sensors	Spatial Resolution	Temporal Resolution	Acquisition and Operational Cost	Coverage Area	Key Limitations	Papers
Smartphone + DL	Variable (10-50 cm)	User-defined (continuous)	Very low; leverages existing consumer devices with negligible ongoing cost	Paddock-scale (manual sampling)	Requires standardisation protocols, lighting variability, limited automation	(McCarthy et al., 2022)
Ground-based Proximal Sensors	Point measurements	Continuous (automated) or manual	Low to moderate equipment cost; labour-intensive operation adding ~AUD$21–38 ha^{-1} $year^{-1}$	Point or transect	Spatial coverage limitations, labour for manual systems	(Lawson et al., 2022; Villalobos-Villalobos & WingChing-Jones, 2023)

Key: SfM = Structure from Motion (photogrammetry for canopy height models); DL = Deep Learning.

Note: Satellite platforms provide consistent temporal monitoring but face spectral saturation above ~3,000 kg DM ha^{-1}. UAV systems enable higher resolution but require operational expertise and favourable weather. Smartphone approaches offer flexibility and low cost but require validation across diverse conditions.

Validation Methodologies and Dataset Limitations

The scarcity of comprehensive, high-quality ground truth datasets represents a pervasive constraint limiting development and validation of robust, transferable prediction models across the pasture remote sensing literature. Correa-Luna et al. (2024) independently evaluated minimum training data requirements for a commercial satellite-based decision support tool in Australian dairy pastures ((Pasture.io.); https://pasture.io), showing that prediction accuracy improved substantially as both the number of field measurements (taken with an rising plate meter calibrated against pasture cuts) and the temporal frequency of calibration increased. Performance stabilised only when ground observations were collected regularly (approximately fortnightly) across seasons and years. The labour-intensive nature of traditional ground-truth protocols such as rising plate meter transects or destructive quadrat harvesting restricts both the spatial extent and temporal frequency of reference datasets, particularly when large numbers of paddocks must be monitored repeatedly. Research across pasture systems consistently demonstrates that limited and spatially uneven ground truth data represent a fundamental bottleneck constraining model calibration and validation in both experimental and commercial applications (Bazzo et al., 2023; Dlamini et al., 2025; Gargiulo et al., 2020).

Rigorous validation of PB estimation systems requires careful consideration of appropriate cross-validation strategies that accurately reflect real-world use case scenarios and avoid optimistic performance estimates arising from inappropriate data splitting procedures. Standard random cross-validation approaches prove fundamentally unsuitable for PB datasets exhibiting strong spatial autocorrelation and temporal dependence, producing inflated performance metrics that fail to generalise to new locations or time periods outside training data distributions (Meyer et al., 2018; Roberts et al., 2017). Spatial cross-validation, whereby training and validation sets are separated based on geographic location, provides realistic assessment of model transferability to new farms or regions. Meanwhile, temporal cross-validation with models trained on historical data and validated on subsequent periods evaluates capacity to generalise across seasons and years critical for operational decision support systems requiring multi-year reliability (Correa-Luna et al., 2024; Stumpe et al., 2023).

Assessment of temporal transferability represents a particularly critical validation dimension frequently neglected in published pasture monitoring studies. Models demonstrating strong performance during development using data from single growing seasons commonly exhibit substantial degradation when applied to subsequent years characterised by different climatic conditions, management practices or phenological patterns, necessitating recalibration or retraining to maintain acceptable accuracy (Morse-McNabb et al., 2023). Ogungbuyi et al. (2024) demonstrated that model performance varied substantially across ten Australian dairy farms despite development using pooled training data, with site-specific environmental conditions and management scenarios introducing variability that generic

models failed to fully capture. The establishment of multi-year validation datasets spanning diverse seasonal conditions proves essential for credible assessment of operational system reliability, though limited research funding horizons and publication pressures frequently constrain temporal extent of evaluation periods in academic studies (Correa-Luna et al., 2024; Filippelli et al., 2024; Smith et al., 2023).

Research Priorities and Economic Considerations

The assessment of cost-benefit relationships and return on investment for remote sensing-based PB monitoring systems across diverse farm scales and production systems remains limited. Quantitative evidence from multi-year farm studies demonstrates substantial economic benefits of improved pasture management in P-BDS. These benefits derive primarily from maximising utilisation of low-cost grazed pasture relative to purchased supplements, and extending grazing season length (Fig. 2.2). However, realising these benefits requires accurate PB information, with traditional measurement approaches costing ~AUD$21-38 ha^{-1} $year^{-1}$ (Rennie et al., 2010; Sanderson et al., 2001). These economic benefits substantially exceed the cost of pasture measurement, providing strong economic justification for investment in monitoring technologies that reduce labour requirements whilst maintaining measurement accuracy.

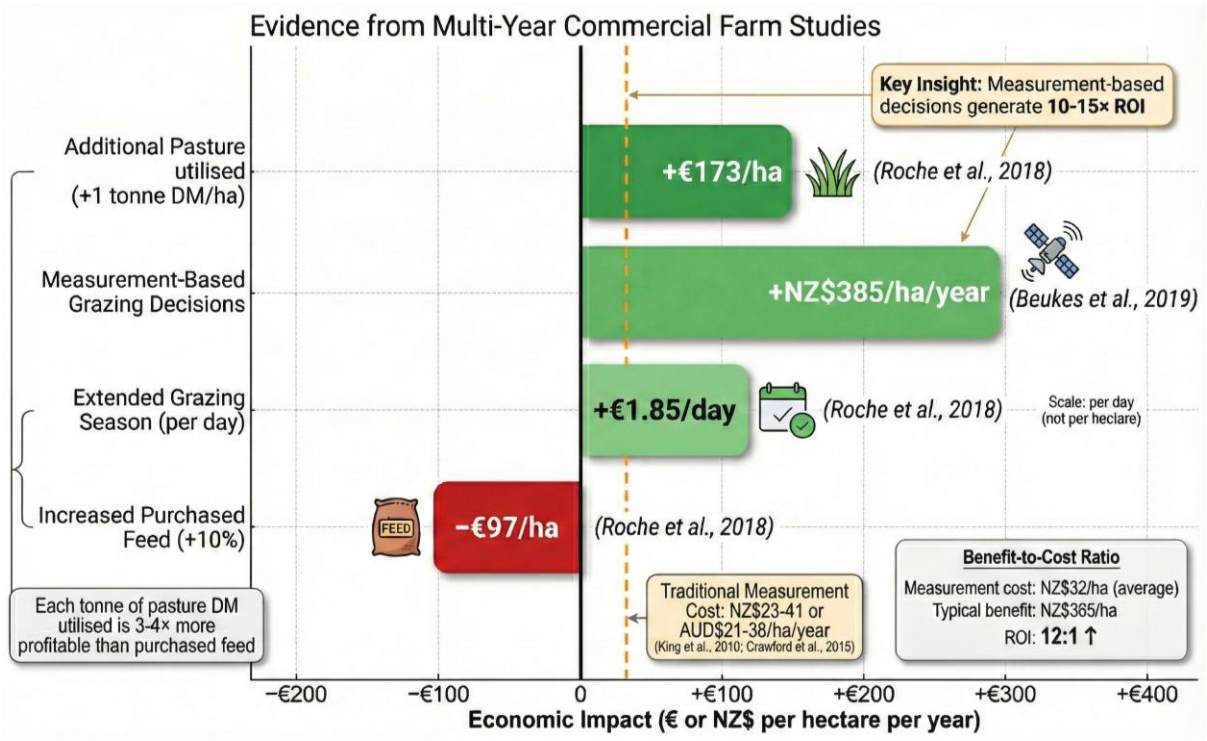


Fig. 2.2 Economic benefits of improved pasture management and monitoring in commercial dairy systems.

Quantified economic impacts from multi-year farm studies in Ireland (8-year dataset) and New Zealand. Green bars show positive benefits from improved management practices; red bar indicates costs from increased

purchased feed. Dashed vertical line represents typical measurement costs, demonstrating that management benefits substantially exceed monitoring expenses

Comprehensive economic analyses incorporating technology costs (including ground truth measurement equipment, processing software, technical support), labour savings relative to traditional monitoring approaches and production benefits from improved feed budgeting and grazing management would provide evidence base for investment decisions. Different monitoring strategies (satellite-only, satellite plus UAV, satellite plus ground sensors) entail different cost structures and performance characteristics, requiring systematic comparison to identify optimal approaches for different operational contexts and management objectives (Cockburn, 2020; De Vries et al., 2023). The integration of PB estimation outputs with whole-farm decision support systems requires development of user-friendly interfaces, visualisation tools and decision algorithms that translate biomass maps into actionable grazing management recommendations. Most farmers lack technical expertise to interpret complex spatial datasets or quantitative model outputs, necessitating development of intuitive, graphics-based decision support tools that present information in accessible formats. Incorporation of economic optimisation frameworks that integrate PB availability with feed costs, milk production responses and financial objectives would enhance value proposition and adoption likelihood of remote sensing-based monitoring systems, though developing such frameworks is non-trivial given the complexity of whole-farm financial objectives, long-term planning horizons, and farm-specific risk considerations that extend well beyond pasture biomass and commodity prices.

In summary, critical research gaps constraining commercial application include the incomplete resolution of spectral saturation effects at high PB levels where tactical grazing decisions are most critical, the absence of robust temporal interpolation methods addressing cloud-induced data gaps in high-rainfall climates, and the limited availability of multi-scenario validation datasets enabling assessment of cross-seasonal performance stability. The pathway towards widespread adoption requires coordinated efforts across near-term and long-term priorities. In the near term, establishing standardised ground truth collection protocols reducing labour intensity whilst maintaining spatial and temporal coverage, and developing transfer learning frameworks enabling model adaptation across regions with minimal local calibration data represent achievable objectives. Longer-term priorities include methodological advances in multi-sensor fusion algorithms integrating optical, SAR, weather data and proximal sensing data streams, and comprehensive economic analyses quantifying return on investment across diverse farm scales and production systems. Resolving spectral saturation and cloud gap limitations would enable continuous automated biomass monitoring to replace labour-intensive manual measurement, substantially reducing operational costs whilst improving the timeliness and spatial coverage of grazing management decisions. Accurate, continuous PB estimation forms the essential foundation for subsequent components of precision grazing management, enabling informed paddock

allocation decisions and providing the baseline feed availability data necessary for calculating optimal supplementation strategies when integrated with grazing behaviour monitoring and individual cow feeding requirements.

GRAZING MANAGEMENT AND PASTURE UTILISATION MONITORING

Accurate estimation of available PB, whilst fundamental to feed budgeting decisions, represents only one component of the information required for precision management in P-BDS. Knowing how much pasture is available before grazing provides limited operational value without corresponding knowledge of how much was actually consumed during grazing events (GE) and what residual PB remains for regrowth. This distinction between pre-grazing PB availability and post-grazing residual levels defines pasture utilisation efficiency, a critical metric that directly influences subsequent pasture regrowth rates, nutritive value development and whole-farm feed budgeting accuracy (Gaffney et al., 2018; Holtgrave et al., 2023). The capacity to quantify utilisation efficiency across multiple paddocks and grazing cycles enables identification of suboptimal management practices where excessive residuals indicate underutilisation whilst insufficient residuals compromise pasture persistence and productivity. However, current approaches for monitoring GEs and calculating utilisation patterns rely predominantly on labour-intensive manual observations or expensive wearable sensor technologies. This creates a fundamental gap in the precision management toolkit that prevents operational implementation of closed-loop systems linking pasture availability with actual consumption patterns (Chebli et al., 2022; Lamanna, Bovo, & Cavallini, 2025; Nyamuryekung'e et al., 2023).

Pasture Utilisation Efficiency and Management Decision-Making

Pasture utilisation efficiency is one of the key factors determining the economic sustainability of grazing operations, with broader implications for environmental outcomes through its influence on supplementary feed demand and long-term pasture management. It governs the balance between maximising forage consumption to reduce supplementary feed costs and maintaining adequate residuals for pasture persistence and productivity (García et al., 2014). Optimal utilisation balances competing objectives of maximising forage consumption to reduce supplementary feed requirements whilst maintaining sufficient residual PB to sustain rapid regrowth and prevent pasture degradation. The relationship between grazing intensity and subsequent pasture productivity exhibits strong non-linear characteristics, with both undergrazing and overgrazing compromising long-term pasture productivity through different mechanisms (García et al., 2014). Insufficient defoliation results in excessive accumulation of mature, low-quality herbage that shades emerging tillers and reduces overall pasture quality, whilst severe defoliation depletes carbohydrate reserves and compromises plant vigour,

particularly under environmental stress conditions or when combined with inadequate rest periods between GEs (Greenwood et al., 2018).

Quantifying actual utilisation efficiency requires accurate measurement of both pre-grazing and post-grazing PB levels across individual paddocks and grazing events. The difference between these measurements indicates the quantity of pasture consumed during the grazing period, enabling calculation of pasture utilisation percentage relative to available PB. This information is essential for multiple management decisions including adjustment of stocking rates to match feed availability, optimisation of rotation schedules to ensure adequate regrowth periods, and calculation of supplementary feed requirements when pasture consumption falls short of animal nutritional needs. Furthermore, systematic monitoring of utilisation patterns across seasons and years provides diagnostic information regarding grazing behaviour, paddock preferences and the effectiveness of rotational grazing strategies. Utilisation patterns can also reveal underlying agronomic issues including inadequate fertiliser application, weed encroachment (Narayanan et al., 2021; U et al., 2022), or poor pasture species establishment that constrain biomass production and forage quality (García et al., 2014). Paddocks exhibiting consistently low utilisation may indicate issues with pasture quality, water accessibility or fence configurations that constrain animal movement. In contrast, paddocks showing excessive utilisation suggest allocation errors where insufficient area was provided relative to herd size and feed demand (Lobert et al., 2021; Nickmilder et al., 2021).

Current Pasture Utilisation Monitoring Approaches and Operational Constraints

Visual observation and manual recording of grazing activities is the most widely adopted approach for tracking animal location and inferring pasture utilisation patterns in commercial dairy operations. Farmers typically rely on visual observation during routine farm activities to assess which paddocks contain livestock at different times, recording this information in paper-based or digital farm management systems (Chapman et al., 2016). This approach requires minimal capital investment and leverages existing farm knowledge and observation skills. However, the temporal and spatial limitations of manual observation methods constrain capacity to detect subtle changes in grazing behaviour or utilisation efficiency that may indicate emerging management issues. Observations provide only snapshot assessments at specific time points rather than continuous monitoring, potentially missing short-duration GEs or changes in animal distribution within paddocks. The labour demands of systematic observation across large farms managing numerous paddocks is prohibitive, whilst subjective variability in visual PB assessment introduces uncertainty into pre-grazing and post-grazing PB estimates used for utilisation calculations (Yang et al., 2011).

Wearable sensor technologies have been developed to address limitations of manual observation through continuous, automated tracking of animal location and behaviour patterns. Global Positioning System (GPS) collar systems deployed on individual animals record position coordinates at regular intervals ranging from seconds to minutes, enabling reconstruction of spatial movement patterns and inference of paddock occupancy throughout each day. Integration of GPS location data with tri-axial accelerometers enables classification of behaviour states including grazing, ruminating, resting and walking based on characteristic head and body movement patterns. These systems provide detailed time budgets and activity profiles for individual animals or grazing mobs (Vázquez Diosdado et al., 2015). Chebli et al. (2022) investigated GPS collars and sensors for monitoring grazing behaviour and energy balance of goats browsing in Mediterranean forest rangeland, deploying GPS collars recording position at 5-minute intervals combined with tri-axial accelerometers capturing head movement patterns. Classification algorithms applied to accelerometer data achieved accuracy exceeding 85% for distinguishing grazing from non-grazing activities, enabling quantification of daily grazing duration and spatial distribution patterns across heterogeneous rangeland landscapes (Chebli et al., 2022). In the same vein, Nyamuryekung'e et al. (2023) demonstrated real-time monitoring of grazing cattle using Long Range Wide Area Network (LoRaWAN) sensors to detect animal welfare implications via daily distance walked metrics, with GPS collars transmitting location data at hourly intervals to cloud-based processing platforms in extensive rangeland systems.

Despite these demonstrated capabilities for automated behaviour monitoring highlighted above, several practical constraints prevent widespread adoption of GPS collar technologies in commercial P-BDS. First, capital costs averaging several hundred dollars per collar create substantial investment requirements for large herds, with a ~350-cow dairy (current average of the Australian dairy industry, Dairy Australia In Focus 2025) requiring expenditure exceeding tens of thousands of dollars for whole-herd monitoring. Second, battery life constraints necessitate periodic collar removal for recharging, creating labour demands and potential data gaps during recharging periods. Third, collar durability under harsh environmental conditions including rain, mud and animal interactions affects equipment longevity and replacement costs. Fourth, data transmission challenges in areas with limited cellular or wireless network coverage may necessitate deployment of supplementary infrastructure including base stations or repeaters, further increasing system costs and complexity (Aquilani et al., 2022; Chebli et al., 2022; Lamanna, Bovo, Bellisola, et al., 2025; Lamanna, Bovo, & Cavallini, 2025; Nyamuryekung'e et al., 2023). Emerging virtual fencing systems such as Halter (New Zealand) and eShepherd (Australia) integrate GPS tracking, accelerometer-based behaviour monitoring and remote paddock allocation control through collar-mounted devices, addressing several traditional GPS collar constraints through automated grazing management (Anderson, 2007; Herlin et al., 2021; Verdon et al., 2024; Verdon et al., 2021). However, whole-herd deployment costs, ongoing connectivity subscription fees and logistical complexity of collar management continue to limit adoption, maintaining demand for

alternative low-cost monitoring approaches such as satellite-based detection methods. Ear-tag accelerometer sensors represent an emerging lower-cost alternative to GPS collar systems, with longitudinal behavioural data from these devices increasingly supporting management decisions beyond simple activity monitoring (Cavallini et al., 2025).

Satellite-Based Detection Potential and Temporal Resolution Constraints

Satellite-based approaches offer potential for low-cost, non-invasive monitoring of grazing activities through detection of vegetation change patterns resulting from animal defoliation. The temporal signature of vegetation indices following GEs provides indirect evidence of livestock presence and consumption intensity through detection of rapid PB decreases inconsistent with gradual seasonal senescence or meteorological stress patterns. Gaffney et al. (2018) demonstrated strong correlations between stocking rate and remotely-sensed PB metrics at pasture scale, with fall mean PB (correlation coefficient of -0.52 to -0.56) and the 10th percentile of relative difference between summer and fall PB (correlation coefficient of -0.47 to -0.52) providing robust indicators of grazing intensity in bunchgrass-dominated systems. These relationships enable inference of management practices and pasture utilisation patterns from satellite time series without requiring direct animal monitoring through GPS collars or behavioural sensors, substantially reducing infrastructure and operational costs whilst enabling retrospective analysis and benchmarking across multiple farms or regions.

However, distinguishing grazing-induced vegetation change from confounding factors including natural senescence, drought stress, disease or management operations presents the primary methodological challenge for automated GE detection systems. The magnitude of vegetation index response to grazing depends critically on pre-grazing PB level, grazing intensity quantified as proportion of available PB consumed, and spectral characteristics of residual vegetation following defoliation (Gaffney et al., 2018; Numata et al., 2008). At high pre-grazing PB levels exceeding 3 t DM ha^{-1}, vegetation indices may exhibit minimal response to moderate grazing intensity due to spectral saturation effects previously discussed, whilst at low pre-grazing PB natural variability in vegetation indices may exceed the signal attributable to grazing. The temporal resolution of freely available satellite imagery is insufficient to capture GEs occurring between image acquisition dates, particularly problematic for short-duration rotational grazing where animals occupy individual paddocks for only 1 to 3 days. The five-day nominal revisit of Sentinel-2 for example becomes substantially longer under persistent cloud cover conditions characteristic of temperate coastal and high-rainfall climates where operational applications require GE detection (Holtgrave et al., 2023; Lobert et al., 2021).

Cloud cover persistence can also result in prolonged gaps spanning multiple weeks without usable satellite data, preventing detection of individual GEs and limiting temporal precision of grazing date

estimates. This temporal discontinuity is particularly problematic in intensive rotational grazing systems common in commercial dairy operations. Multiple paddocks may be grazed in rapid succession, and accurate GE timing is essential for calculating rest periods and planning subsequent grazing allocations. Whilst the temporal constraints described above apply primarily to freely available satellite platforms including Sentinel-2 and Landsat, commercial high-frequency satellite constellations including Planet CubeSat provide daily 3-5 m resolution imagery that substantially alleviates cloud-related data gaps and enables detection of short-duration grazing events (Ogungbuyi et al., 2024). However, commercial satellite data costs create substantial ongoing operational expenses, with whole-farm monitoring potentially exceeding several thousand dollars annually depending on acquisition frequency and spatial coverage requirements, raising questions regarding economic viability for typical commercial dairy systems. The economic trade-off between freely available platforms with limited temporal resolution versus commercial systems providing daily coverage but requiring significant capital investment remains inadequately addressed in published literature.

ML classification approaches trained on datasets combining satellite imagery time series with ground truth records of GEs can learn to distinguish characteristic temporal signatures of grazing from other vegetation change processes, but require substantial training data spanning diverse conditions to achieve robust performance. The scarcity of publicly available datasets containing precisely georeferenced and temporally resolved grazing records limits development and validation of automated GE detection algorithms. Most published studies rely on small datasets from single farms or regions, constraining assessment of model transferability across diverse environmental and management contexts (Correa-Luna et al., 2024; Holtgrave et al., 2023; Lobert et al., 2021).

Multi-Sensor Fusion and Machine Learning for Grazing Management

Integration of optical and synthetic aperture radar time series substantially improves GE detection reliability through complementary sensitivity to vegetation structure and reduced vulnerability to cloud-related data gaps. Holtgrave et al. (2023) demonstrated that fusion of Sentinel-1 SAR with Sentinel-2 optical and meteorological time series enhanced mowing event detection accuracy in permanent grasslands compared to optical-only approaches SAR backscatter provided vegetation structure information independent of atmospheric conditions, whilst meteorological data enabled discrimination of management events from natural vegetation responses to environmental variation. Support vector machine and convolutional neural network classifiers integrating multiple sensor modalities achieved detection accuracies exceeding 85% across diverse grassland management intensities, though performance degraded substantially when models trained on one region were applied to geographically distant locations without local recalibration, highlighting challenges in developing transferable detection algorithms. These findings suggest that single-sensor optical approaches are insufficient for

reliable grazing event detection under cloud-affected conditions, with multi-sensor integration combining optical, SAR, and meteorological data offering the most promising pathway toward operational accuracy.

Lobert et al. (2021) conducted systematic evaluation of input features from Sentinel-1, Sentinel-2 and Landsat 8 time series for mowing event detection, and reported that dense temporal sampling proved critical for distinguishing rapid vegetation change from gradual phenological development or environmental stress responses. The study emphasised that temporal resolution requirements for reliable grazing event detection exceed capabilities of freely available optical satellite platforms under typical cloud cover conditions, necessitating strategic integration of multiple platforms or interpolation approaches to achieve operational monitoring frequency. Integration of contextual information regarding farm management calendars, seasonal patterns and environmental conditions through supervised ML classification approaches enables learning of characteristic temporal signatures that distinguish grazing from alternative vegetation change processes. However, the labour-intensive nature of creating training datasets with accurately labelled grazing events across diverse farms, seasons and management systems constrains development of robust, generalisable detection algorithms (Nickmilder et al., 2021). Table 2.3 summarises the key methods, data, and performance metrics reported across automated grazing detection and pasture utilisation monitoring studies reviewed in this section.

Table 2.3 Summary of automated approaches for grazing event detection and pasture utilisation monitoring.

Study	Method/Sensor	Task	Data	Key Metric
Gaffney et al. (2018)	Satellite biomass metrics	Grazing intensity inference	Semi-arid rangeland, USA	Correlation coefficient −0.52 to −0.56 with stocking rate
Chebli et al. (2022)	GPS collars and tri-axial accelerometers	Grazing behaviour classification	Goats, Mediterranean forest rangeland	>85% accuracy distinguishing grazing from non-grazing
Nyamuryekung'e et al. (2023)	LoRaWAN GPS sensors	Real-time grazing monitoring	Cattle, extensive rangeland systems	Hourly location transmission to cloud platform
Holtgrave et al. (2023)	SVM and CNN with optical, SAR and meteorological data	Mowing event detection	Permanent grasslands, Germany	>85% detection accuracy
Lobert et al. (2021)	Sentinel-1, Sentinel-2 and Landsat 8 time series	Mowing event detection	Permanent grasslands, Germany	Dense temporal sampling critical for reliable detection
Nickmilder et al. (2021)	ML with Sentinel-1, Sentinel-2 and meteorological data	Compressed sward height prediction	Walloon pastures, Belgium	Multiple data transformations evaluated

Note: SVM = support vector machine; CNN = convolutional neural network; SAR = synthetic aperture radar; GPS = global positioning system; LoRaWAN = long range wide area network. Detection accuracies and performance metrics reflect different tasks, sensors, animal species, and environmental contexts and are not directly comparable across studies.

Research Priorities and Pathways to Operational Precision Grazing Systems

A fundamental challenge constraining the implementation of automated grazing event detection is the insufficient temporal resolution of freely available satellite platforms for reliably detecting short-duration GEs characteristic of intensive rotational systems. The associated research gap therefore lies in developing methods that work effectively within these temporal constraints (Holtgrave et al., 2023; Lobert et al., 2021). Whilst satellite-based vegetation monitoring provides paddock-scale spatial coverage at substantially lower cost than GPS collar technologies, the temporal sampling frequency is inadequate for capturing grazing events lasting approximately 1 to 3 days, particularly when compounded by cloud-induced data gaps. This temporal resolution limitation prevents reliable calculation of pre-grazing and post-grazing PB levels that are essential for quantifying utilisation

efficiency. Without automated detection of grazing event timing, integration of PB estimation models developed through approaches with utilisation monitoring remains impractical. Manual recording of grazing dates reintroduces human dependency and the potential for inconsistent or incomplete records across multiple paddocks simultaneously, undermining the continuity and reliability that automated monitoring aims to provide (Heins et al., 2023).

Additionally, development of automated grazing detection systems represents a critical enabling technology for precision grazing management, with cost-effectiveness determining commercial viability alongside accuracy requirements. Without automated detection, manual recording reintroduces labour costs whilst GPS collar technologies achieving high accuracy face capital costs that constrain whole-herd adoption (Chebli et al., 2022; Lamanna, Bovo, & Cavallini, 2025). The integration framework requires three interconnected components operating in concert. First, continuous PB estimation using satellite or ground-based computer vision approaches to provide pre-grazing PB assessment informing tactical paddock allocation decisions. Second, automated detection of GE timing to enable calculation of post-grazing residual PB through comparison of vegetation indices immediately before and after livestock presence in each paddock. Third, quantification of pasture intake through differences between pre-grazing and post-grazing PB can provide the baseline for calculating supplementary feed requirements.

Critical research priorities for advancing automated grazing detection span near-term and longer-term objectives. In the near term, development of temporal interpolation methods combining optical and SAR satellite data with meteorological information would increase effective temporal sampling frequency under cloud-covered conditions, whilst establishment of comprehensive multi-farm training datasets containing precisely georeferenced grazing records across diverse environmental conditions, management systems and seasonal patterns would enable robust ML classification algorithms with demonstrated transferability. Longer-term, investigation of high-frequency nanosatellite constellations such as Planet CubeSat presents opportunities to increase temporal coverage and reduce cloud-related data gaps compared to Sentinel-2, though economic feasibility at commercial farm scale requires systematic evaluation given substantially higher data costs relative to freely available platforms (Gargiulo et al., 2020). Systematic comparison of alternative sensor fusion frameworks and ML architectures would identify optimal approaches for distinguishing grazing-induced vegetation change from confounding processes including senescence, drought stress and management operations (Correa-Luna et al., 2024; Shahi et al., 2025). Cost-effective automated grazing detection would eliminate dependence on expensive wearable sensor technologies, making utilisation efficiency monitoring accessible to commercial farms at scale.

ARTIFICIAL INTELLIGENCE FOR SUPPLEMENTARY FEED ALLOCATION IN PASTURE-BASED DAIRY SYSTEMS

Quantifying PB availability through satellite RS and ML algorithms provides essential information for tactical grazing allocation decisions, whilst automated detection of grazing events enables calculation of actual pasture consumption and utilisation efficiency. However, within the integrated precision management context of this review, the full operational value of these monitoring capabilities is only realised when combined with corresponding systems that translate pasture availability and pasture utilisation patterns into individual cow feeding strategies. Pasture monitoring nonetheless retains independent value for rotation planning and stocking rate decisions regardless of supplementation level. The integration loop remains incomplete until real-time PB estimates and consumption data inform precision supplementation decisions that account for the deficit between pasture intake and individual animal requirements. Pasture-based dairy systems face an economic challenge where pasture alone provides insufficient energy and nutrients to support high milk production levels, necessitating concentrate supplementation to bridge the nutritional deficit. This supplementation creates economic tension because concentrate is typically 3 to 5 times more expensive per kilogram of dry matter than pasture (Clark et al., 2018; Kolver, 2003). Supplementary concentrate represents between 40-60% of total production costs in P-BDS, creating tension between production potential and feed costs that demands precise determination of optimal supplementation levels for each cow (Craig et al., 2022).

The critical question becomes: how much concentrate should each individual cow receive to maximise profit given her unique capacity to convert supplementary feed into milk production whilst accounting for variable pasture availability and quality? Despite this reality, approximately 78% of Australian and New Zealand dairy operations continue to employ flat-rate feeding strategies that distribute identical concentrate allowances across all cows regardless of individual production potential, physiological state, or nutritional requirements (Akintan et al., 2025; Dela Rue & Eastwood, 2017). This persistent reliance on simplified herd-average approaches reflects not complacency but rather the absence of practical, validated systems capable of automating individualised feeding decisions in commercial grazing environments. The documented potential for precision feeding to increase milk solids yield by approximately 7% compared to flat-rate allocation (García et al., 2007) establishes compelling economic incentive for technological innovation, creating opportunity for AI and ML to transform feed management from crude averaging toward genuine optimisation of individual animal responses that completes the precision management loop linking pasture availability, utilisation efficiency, and cow-specific supplementation requirements.

Biological Variability and the Inadequacy of Linear Prediction Models

Individual dairy cows within nominally homogeneous herds exhibit substantial variation in MY responses to supplementary feeding arising from genetic differences, metabolic efficiency, stage of lactation, body condition, health status, and behavioural characteristics that traditional herd-level management approaches cannot accommodate (Hills et al., 2015). This biological heterogeneity manifests as non-linear dose-response relationships where the marginal MY increase per kilogram of concentrate diminishes as supplementation levels rise, with response curves differing substantially between individual animals (Liebe & White, 2019). Flat-rate feeding strategies that ignore this variability inevitably overfeed some cows whose marginal response to additional concentrate has reached minimal levels whilst simultaneously underfeeding others capable of converting supplementary nutrients into additional milk production at economically favourable rates. The aggregate consequence of this mismatch between supply and individual requirements creates systematic inefficiency where total feed costs exceed minimum levels necessary to achieve observed herd MY whilst simultaneously constraining potential production from cows capable of higher performance (Clark et al., 2018).

Traditional biological models attempting to predict individual cow nutritional requirements and MY responses employ linear or logistic regression frameworks incorporating variables including liveweight, days in milk, parity, and recent production history (Krizsan et al., 2014; Raedts & Hills, 2024). These parametric approaches assume that relationships between predictor variables and MY responses conform to predetermined functional forms specified a priori by model developers, constraining capacity to capture complex interactions and non-linear responses characteristic of biological systems (Little et al., 2016). Evaluation of five widely employed feed intake prediction models using 2,161 individual cow observations across 24 trials in Nordic countries demonstrated mean prediction errors ranging from 1.62 to 3.19 kg DM d^{-1} for dry matter intake (DMI), with substantial variation in accuracy across different environmental conditions and management scenarios (Krizsan et al., 2014). This magnitude of prediction error, when propagated through feeding decisions across entire lactations and multiple animals, translates into considerable economic losses from suboptimal resource allocation. he fundamental limitation lies not in insufficient model complexity but rather in the parametric assumption that biological responses can be adequately characterised through linear combinations of predictor variables, particularly when complex interactions and non-linear relationships are present among cows (Raedts & Hills, 2024).

Machine Learning for Capturing Non-Linear Individual Cow Variability

ML algorithms offer methodological alternatives to traditional linear models through capacity to learn complex, non-linear relationships directly from data without requiring explicit specification of functional forms. Tree-based ensemble methods such as Random Forest and Gradient Boosting

Machines construct predictions by combining multiple weak learners, enabling capture of interactions and non-linearities that confound parametric approaches (Shine & Murphy, 2022). Neural networks learn hierarchical representations through multiple processing layers, with deep architectures demonstrating particular capability for extracting subtle patterns from high-dimensional datasets incorporating diverse data streams (Martin et al., 2021). These data-driven approaches are especially valuable for modelling individual cow MY responses where relationships between predictor variables and outcomes exhibit the complexity characteristic of biological systems rather than conforming to predetermined mathematical functions.

Recent implementations of ML for individual cow feeding decisions demonstrate superior predictive performance compared to conventional biological models. Kamphuis et al. (2017) applied boosted regression trees to predict individual feed intake in dairy cows from a single research farm in the Netherlands using 245 cows for model development and 155 cows for validation, achieving moderate correlations ($r = 0.73$) between actual and predicted intakes using feeding behaviour patterns, demonstrating viability of behaviour-based predictions in commercial environments. Davison et al. (2021) employed linear mixed models incorporating feeding behaviour variables to predict DMI in finishing beef steers across 80 animals over a 56-day period in Scotland, attaining R^2 values between 0.59 and 0.73 across different analytical techniques, confirming that behavioural patterns captured through automated monitoring systems contain information content sufficient for intake prediction. For dairy-specific applications, Song et al. (2025) developed individual cow models for predicting daily MY using nonlinear regression analysis applied to 580 individual cow datasets, achieving $R^2 = 0.875$ and RMSE = 2.2 kg, enabling real-time anomaly detection for health monitoring alongside feeding optimisation.

Ensemble methods that integrate multiple data streams demonstrate enhanced performance through synergistic combination of complementary information sources. Martin et al. (2021) compared multiple linear regression, partial least squares regression, artificial neural networks, and stacked ensembles using combinations of cow descriptive data, performance metrics, sensor-derived behavioural information, and blood metabolite measurements from 124 mid-lactation Holstein cows. Stacked ensemble approaches consistently outperformed single-algorithm methods for predicting DMI and residual feed intake. This demonstrated that integration of diverse data sources through sophisticated ML architectures enables more accurate characterisation of individual cow requirements than achievable through any single data stream or analytical method. Similarly, Monteiro et al. (2024) leveraged ensemble methods with feature engineering to explore rumen microbiome contributions to feed efficiency in 454 genotyped Holstein cows, successfully identifying key microbial taxa linked to residual feed intake whilst demonstrating that rumen microbiome composition represents a major driver of between-animal variation in feed conversion efficiency.

Computer vision approaches using red-green-blue-depth (RGB-D) camera systems integrated with deep DL algorithms enable automated monitoring of individual feeding behaviour without requiring manual intervention or specialised sensing equipment attached to animals. Saar et al. (2022) investigated computer vision applications for measuring individual feed consumption through RGB-D cameras combined with convolutional neural network architectures. Performance evaluation revealed prediction errors of 0.12 to 0.13 kg per feeding event mean absolute error (MAE) and 0.17 to 0.18 kg per feeding event RMSE when validated against manual data collection, with real-world automated testing producing MAE = 0.14 kg per feeding session and RMSE = 0.19 kg per feeding session. These accuracy levels approach requirements for practical farm-scale implementation whilst eliminating labour associated with manual intake measurement or maintenance demands of individual animal sensors. Table 2.4 summarises the key ML approaches, data, and performance metrics reported across individual cow feeding prediction studies reviewed in this section.

Table 2.4 Summary of machine learning approaches for individual cow feeding prediction in pasture-based dairy systems.

Study	Algorithm	Outcome Variable	Data	Key Metric
Kamphuis et al. (2017)	Boosted Regression Trees	Individual feed intake	245 cows for development, 155 for validation, single research farm, Netherlands	r = 0.73
Davison et al. (2021)	Linear Mixed Models	Dry matter intake	80 finishing beef steers, 56-day period, Scotland	R^2 = 0.59–0.73
Martin et al. (2021)	Stacked Ensembles	DMI and residual feed intake	124 mid-lactation Holstein cows, multiple data streams	Outperformed single-algorithm methods
Monteiro et al. (2024)	Ensemble Methods with feature engineering	Feed efficiency	454 genotyped Holstein cows, rumen microbiome data	Identified key microbial taxa linked to RFI
Song et al. (2025)	Nonlinear Regression	Daily milk yield	580 individual cow datasets	R^2 = 0.875, RMSE = 2.2 kg
Saar et al. (2022)	Convolutional Neural Network	Individual feed consumption per event	RGB-D camera system, automated farm setting	MAE = 0.12–0.14 kg per event

Note: R^2 = coefficient of determination; r = Pearson correlation coefficient; MAE = mean absolute error; RMSE = root mean square error; DMI = dry matter intake; RFI = residual feed intake. Metrics reflect different outcome variables, animal species, and study contexts and are not directly comparable across studies

Optimisation Algorithms for Individualised Feed Allocation

Predicting individual cow MY responses to varying concentrate levels, whilst necessary, proves insufficient for operational feeding decisions without corresponding optimisation frameworks that translate predictions into actionable allocation strategies aligned with farm objectives and constraints. Linear programming provides the most widely implemented mathematical approach for feed formulation, balancing nutritional requirements against economic constraints through optimisation of linear objective functions subject to linear equality and inequality constraints (Bellingeri et al., 2020). Implementation of linear programming for 29 Holstein-Friesian herds with average size of 313.2 ± 144.1 lactating cows demonstrated capacity to formulate diets for each animal group whilst respecting DMI limits and fulfilling crude protein, energy, and fibre requirements, achieving substantial cost reductions compared to conventional formulation approaches. The Simplex method of linear programming has been deployed in practical web-based tools including OPTIMILK, which accumulated 62 registered users who saved 581 rations over a 3.5-year evaluation period, confirming commercial viability of mathematical programming approaches for routine farm management (Mijić et al., 2024).

Nonlinear programming extends optimisation capabilities beyond linear assumptions to address inherent nonlinearities in animal nutrition responses. Li et al. (2022) applied non-linear programming to dairy cattle ration formulation for 841 cow-period observations across 31 chamber studies, incorporating gross energy intake variations (coefficient of variation = 0.10) and milk energy output efficiency (coefficient of variation = 0.084) to optimise feeding strategies. The nonlinear approach demonstrated superior performance compared to linear models when addressing complex interactions between nutrients and animal responses, particularly for handling between-cow variations in feed efficiency components. Stochastic programming models extend deterministic optimisation frameworks by incorporating uncertainty in feed composition and market prices, addressing real-world variability affecting dairy operations through probabilistic approaches that prove particularly valuable for P-BDS where feed quality varies substantially with environmental conditions (Dowson et al., 2019; Peña et al., 2009).

Evolutionary algorithms including genetic algorithms and differential evolution enable multi-objective optimisation that addresses competing goals beyond simple cost minimisation through population-based search strategies that identify trade-offs amongst multiple objectives. Although more complex than other optimisation approaches like linear programming, evolutionary algorithms have the advantage of reduced susceptibility to becoming trapped in local optima through their population-based parallel search mechanisms (Cao et al., 2016; Peña et al., 2009). Notte et al. (2020) implemented multi-objective optimisation models using differential evolution algorithms for dairy feeding management,

evaluating mono-objective and multi-objective scenarios across stocking rates ranging from 1.1 to 2.6 cows ha^{-1} using 48 Holstein cows in a 17-week randomised complete block design. The differential evolution solutions approached linear programming optimal solutions with differences between 0.23% and 6.17% across stocking rates, whilst analysis demonstrated that increasing stocking density enhanced MY and gross margin per unit area whilst shifting feed allocation from roughage toward concentrate supplementation. Extension to multi-objective frameworks using both genetic algorithms and differential evolution enables generation of Pareto front approximations that present decision-makers with sets of non-dominated solutions representing different trade-offs between objectives including profitability maximisation, feed cost minimisation, and environmental impact reduction (Notte et al., 2021).

The integration of ML prediction models as surrogate models within optimisation frameworks represents a critical technical opportunity for precision feeding implementation. Conventional optimisation approaches employing mechanistic biological models face computational bottlenecks when evaluating large solution spaces across multiple animals and time periods, particularly for nonlinear and stochastic formulations requiring iterative solver algorithms. Highly performant ML models trained on comprehensive datasets can generate MY predictions orders of magnitude faster than mechanistic simulations whilst maintaining or exceeding prediction accuracy, enabling their use as surrogate models within optimisation loops that require thousands of function evaluations (Franzoi et al., 2022; Wang et al., 2020; Williams & Cremaschi, 2021). This surrogate modelling approach can be particularly valuable for evolutionary algorithms including genetic algorithms and differential evolution that operate through iterative improvement of candidate solutions, where replacing mechanistic model evaluations with rapid ML predictions dramatically reduces computational time whilst maintaining optimisation quality. Single-objective optimisation frameworks focus on maximising profit or minimising cost subject to nutritional and operational constraints, whilst multi-objective approaches simultaneously address competing goals enabling identification of feeding strategies that balance economic returns with environmental sustainability metrics or animal welfare considerations.

The Critical Gap: Automated Integration and Demonstrated Performance in P-BDS

The fundamental research gap constraining practical adoption of AI-driven precision feeding lies in the absence of empirical evidence demonstrating that automated systems integrating multiple data sources with ML and optimisation algorithms can actually deliver increased milk yield by leveraging biological variability amongst individual cows to determine optimal individualised supplementation levels. Whilst ML algorithms demonstrate superior predictive performance for individual cow responses compared to linear models (Kamphuis et al., 2017; Song et al., 2025), and optimisation frameworks including genetic algorithms and differential evolution can identify optimal feeding strategies, no published research has

experimentally validated whether fully automated systems translating these predictions into practice produce measurable MY improvements and cost savings in commercial P-BDS. The 7% milk solids increase from individualised feeding documented by García et al. (2007) was achieved through manual adjustment based on predetermined rules rather than AI-driven automation, leaving unresolved whether automated ML-optimised systems can replicate or exceed these gains whilst simultaneously reducing labour demands (Clark et al., 2018; Craig et al., 2022). Beyond algorithmic performance, this validation gap carries direct economic consequences, as prediction errors propagating through automated feeding decisions across entire lactations and large herds translate into quantifiable losses from suboptimal concentrate allocation.

The challenge extends beyond algorithm development to practical integration of heterogeneous data sources that collectively characterise individual cow requirements and environmental constraints. Effective precision feeding demands synthesis of cow-specific characteristics including liveweight, days in milk, parity, genetic merit, body condition score, and recent production history. Real-time information streams including weather patterns affecting pasture quality and intake behaviour, automated milk yield recording from milking systems, behavioural sensor data capturing rumination and activity patterns, and feed intake measurements from electronic feeding stations are equally essential (Kamphuis et al., 2017; Martin et al., 2021; Song et al., 2025). Current ML model development relies predominantly on retrospective observational data reflecting existing flat-rate or narrow-range feeding practices, constraining model capacity to learn dose-response relationships across the full range of physiologically plausible supplementation levels (Dela Rue & Eastwood, 2017). Models trained on such datasets can identify correlations between cow characteristics and production outcomes under observed management, but experimental validation through automated systems that actively manipulate feeding rates based on ML predictions and optimisation algorithms remains absent from the published literature (Souza et al., 2022).

The distinction between demonstrating that ML models can predict individual responses and proving that AI-automated systems leveraging these predictions can actually increase production carries profound practical and economic implications. Farmers require evidence that automated individualised feeding delivers quantifiable MY gains sufficient to justify capital investment in sensing infrastructure, electronic feeding systems, and data management platforms whilst simultaneously reducing labour compared to conventional flat-rate approaches (Cockburn, 2020; De Vries et al., 2023). Addressing this validation gap requires field trials where ML-optimised feeding recommendations integrating multiple data sources are implemented automatically through electronic feeding systems. These trials should span sufficient animals and duration to capture responses under varying environmental conditions and lactation stages, enabling definitive assessment of whether AI-driven precision feeding delivers measurable improvements in practice (Shine & Murphy, 2022; Tzanidakis et al., 2023).

Beyond immediate quantification needs, the research gap extends to understanding mechanisms through which AI systems achieve improvements by exploiting between-cow variability whilst integrating diverse data streams. Farmers adopting precision feeding technologies require transparent explanations regarding why specific cows receive particular allocations, how weather patterns or behavioural changes influence feeding decisions, and how individual feeding decisions aggregate to herd-level production and economic outcomes. The black-box nature of ensemble ML methods and neural networks producing recommendations without interpretable reasoning limits farmer trust in automated suggestions even when predictions demonstrate statistical accuracy (Cabrera & Fadul-Pacheco, 2021). Development of explainable AI methodologies that provide transparent justifications whilst maintaining prediction accuracy would enhance adoption likelihood, but ultimately farmer confidence depends on demonstrated proof that automated systems deliver measurable, sustained improvements in MY and profitability across realistic operational timeframes and varying environmental conditions rather than theoretical predictions based on correlational patterns in historical datasets.

Implementation Challenges and Research Priorities

Translating algorithmic capabilities into operational precision feeding systems faces multiple technical and practical challenges that constrain commercial adoption despite demonstrated potential. Manual adjustment of individual cow feed allocations is prohibitively time-consuming for large herds, whilst automated feeding systems capable of delivering individualised concentrate require substantial capital investment in electronic identification, controlled access feeding stations, and data management infrastructure often exceeding economic viability thresholds particularly for smaller operations (Cockburn, 2020; De Vries et al., 2023). Integration of diverse data streams including automated MY recording, behavioural sensors, PB estimates from satellite or computer vision systems, and weather data into unified decision support platforms demands technical expertise and robust error handling procedures to address missing measurements, sensor failures, and data quality issues frequently unavailable on commercial farms.

Computational efficiency considerations are critical for practical implementation when optimisation algorithms must process problems involving hundreds of individual animals updated multiple times daily. Leveraging ML models as surrogate functions within evolutionary algorithms optimisation loops including genetic algorithms and differential evolution minimises computational demands compared to mechanistic model evaluations, enabling real-time feeding decisions at farm scale (Franzoi et al., 2022; Notte et al., 2021; Uyeh et al., 2019; Williams & Cremaschi, 2021). However, systematic comparison of alternative evolutionary algorithms frameworks for single-objective versus multi-objective formulations, deterministic versus stochastic programming, and hybrid approaches would identify optimal methodologies for different problem scales and constraint structures. Algorithm explainability

represents emerging priority because ensemble methods and neural networks function as black boxes producing recommendations without transparent reasoning, limiting farmer confidence in automated suggestions (Cabrera, 2025; Cabrera & Fadul-Pacheco, 2021). Development of explainable AI methodologies providing interpretable justifications whilst maintaining prediction accuracy would enhance trust and adoption likelihood.

Critical research priorities center on establishing whether AI-automated individualised feeding delivers quantifiable MY increases by leveraging cow-to-cow variability. In the near term, controlled field trials implementing ML-optimised recommendations through automated feeding systems compared against flat-rate controls across sufficient animals, duration, and farm diversity would enable causal assessment of production responses and economic returns under varying pasture availability and climatic conditions. Integration of real-time PB estimates from satellite or computer vision with automated grazing event detection and individual cow predictions would enable dynamic feeding decisions adapting to actual pasture consumption patterns, closing the precision management loop linking biomass availability, utilisation efficiency, and optimised supplementation (Shorten, 2021). The assessment of minimum infrastructure requirements and alternative sensing configurations would identify cost-effective approaches suitable for resource-constrained operations, whilst comprehensive economic analyses incorporating technology costs, labour savings, and sustained production benefits across diverse farm scales remain essential for evidence-based investment decisions. Looking further ahead, longitudinal assessment spanning five or more years would establish realistic expectations regarding system reliability, maintenance demands, and economic returns under varying market and environmental scenarios, addressing fundamental farmer concerns regarding technology robustness and return on investment. Validated automated individualised feeding systems would transform a theoretically documented production gain into a commercially realisable outcome, simultaneously reducing labour demands and improving feed use efficiency. Successful transition from experimental demonstration to widespread commercial adoption ultimately depends on validated proof that automated precision feeding systems deliver measurable improvements where theoretical algorithms meet operational reality.

In summary, the convergence of advancing capabilities in PB estimation through satellite remote sensing and ML, emerging potential for automated grazing event detection enabling pasture utilisation quantification, and demonstrated superiority of ML algorithms for predicting individual cow responses positions precision management as an achievable goal for commercial P-BDS. However, realising integrated systems requires addressing persistent challenges spanning multiple domains. For PB estimation, spectral saturation at high PB levels, cloud-induced temporal gaps, and limited ground truth availability constrain satellite-based approaches (Correa-Luna et al., 2024; Morse-McNabb et al., 2023), whilst computer vision using smartphone imagery with DL offers complementary capabilities

maintaining sensitivity across broader biomass ranges. For grazing management, insufficient temporal resolution prevents reliable detection of short-duration grazing events characteristic of intensive rotational systems, representing the critical missing link between biomass availability and pasture utilisation efficiency quantification.

For precision feeding, the critical gap lies in the absence of evidence demonstrating that AI-automated systems integrating ML predictions with evolutionary algorithms can quantify and deliver actual milk yield increases by leveraging cow-to-cow variability, with little research validating whether automated systems can reproduce the gains documented through manual individualised feeding. Common challenges pervading all domains include data scarcity, inappropriate validation methodologies producing optimistic performance estimates that fail to generalise (Meyer et al., 2018; Roberts et al., 2017), integration complexity across heterogeneous data streams, and economic barriers particularly affecting smaller operations. Advancing from isolated technological demonstrations toward unified platforms requires sustained research investment addressing fundamental methodological limitations whilst developing practical implementation frameworks that translate algorithmic capability into demonstrable value for commercial dairy operations.

CONCLUSION

This review demonstrates that machine learning applications across pasture biomass estimation, grazing management, and precision feeding have shown substantial promise under research conditions, yet demonstrated algorithmic capability remains largely disconnected from validated operational value. This gap between research performance and farm-level benefit represents the defining challenge constraining progress toward integrated precision management in pasture-based dairy systems. Three unresolved limitations collectively prevent progress toward integrated precision management. Spectral saturation and cloud-induced temporal gaps constrain satellite biomass monitoring precisely where tactical decisions are most urgent. Insufficient satellite temporal resolution prevents reliable grazing event detection, severing the link between biomass availability and utilisation efficiency. No published study has validated whether automated feeding systems could deliver quantifiable milk yield gains by exploiting individual cow variability under commercial conditions. Until these limitations are resolved, integration across the three domains remains aspirational rather than operational. Future research should prioritise controlled field validation of automated feeding systems, cost-effective approaches to grazing event detection, and multi-sensor fusion strategies that can address satellite data gaps. If these priorities are pursued systematically, integrated precision management platforms could potentially deliver measurable improvements in feed efficiency, pasture utilisation, and milk production for pasture-based dairy operations globally.

This chapter has been accepted for publication in *Computer and Electronics in Agriculture*, with the title '**AI in Pasture-Based Dairy Systems: Advances in Precision Feeding, Remote Sensing, and Grazing Management**' (Manuscript ID: COMPAG-D-25-09096).

CHAPTER 3

A data-driven approach for optimising supplement allocation to individual lactating dairy cows in pasture-based systems

Every year, most dairy farmers pour their largest expense, feed costs, into systems treating every cow the same. This chapter confronts this head on. Conventional flat rate feeding ignores the fundamental reality that cows respond differently to the same ration. Recognising that traditional models cannot capture cow specific responses, we introduce a data driven framework using nonlinear machine learning and optimisation models to individualise concentrate allocation from existing feeding records. This reveals a quantifiable theoretical potential of an 8 % increase in total milk yield while maintaining the original feed budget, transforming flat rate feeding into precision allocation and enhancing profitability. The complete model comparison analysis and optimisation implementation are available in the repository. The model comparison notebook (modelling.ipynb) evaluates Random Forest, Neural Networks, LSTM, and Gaussian Process models. The optimisation script (scipy_optimisation.py) implements Scipy differential evolution algorithm for day-by-day concentrate allocation over 91 days. Key code excerpts are presented in Appendix A.

Smart Agriculture Technology (2025) 12, 101669

PUBLISHED MANUSCRIPT

The published version of this manuscript is included on the following page.

Smart Agricultural Technology 12 (2025) 101669

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

A data-driven approach for optimising supplement allocation to individual lactating dairy cows in pasture-based systems

Blessing Nnenna Azubuike [a,*], Anna Chlingaryan [b], Martin Correa-Luna [a], Cameron E.F. Clark [b], Sergio C. Garcia [a]

[a] Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW 2570, Australia
[b] Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Camden 2570, Australia

ARTICLE INFO

Keywords:
Differential evolution
Feed optimisation
Individual concentrate allocation
Machine learning
Pasture-based systems

ABSTRACT

With feed costs accounting for about 40-60 % of milk production expenses in Australia, efficient supplementary concentrate allocation is crucial for profitability. Despite an increase in concentrate use per cow over the past decade, the average milk yield response remains about 1 L per kg of dry matter concentrate. While machine learning and data-driven optimisation are widely utilised in sectors such as engineering, healthcare, and finance, their application in feed optimisation within dairy farming has not been extensively researched. This study aims to develop a machine learning-based method to optimise individual cow supplement allocation, using similar total daily concentrate with a tolerance range allowing for a 2-10 % decrease, to maximise milk yield. Data from a controlled field study involving 130 lactating Holstein-Friesians (32,504 records) were analysed.

Sixteen machine learning algorithms were evaluated to predict milk yield based on concentrate allocation and available cattle data (days in milk, daily milk yield and liveweight, and parity number). The Random Forest (RF) model was the best performer, achieving an R^2 of 0.60 and RMSE of 4.20 L/cow/day. Then 7371 records from 81 cows over 91 days were employed to run the concentrate levels optimisation using the Dirichlet-Rescale (DRS) algorithm and Monte Carlo simulation. The RF model and SciPy optimisation determined optimal individual cow allocations (5-9 kg/cow.day^{-1}). Implementing this resulted in a herd-level 8 % increase in daily milk yield. This study highlights the potential benefits of adopting data-driven algorithms for individualised dairy feed optimisation based on observed correlations within existing management practices. While results suggest improvement over conventional flat-rate methods, the study is limited by the nature of the dataset, and findings reflect associations under standard practice rather than experimental manipulation of concentrate levels, requiring validation through controlled field trials to confirm practical efficacy and economic impact in actual dairy farming contexts.

1. Introduction

Pasture-based dairy farms are the predominant dairy production systems in Australia. In these systems, the cost of milk production is inversely related to profitability [1]. Feed accounts for about 40-60 % of the cost of producing milk and grain-based concentrates are typically 3 to 5 times more expensive per kg DM than pasture in Australia [2,3]. Cow nutrition is affected by seasonal variations in both pasture quality and quantity, weather-related fluctuations. To counter this, supplementary feed, including fodder and concentrates are used by dairy farmers to uphold cow health and production during periods of low pasture availability. Over the last 10-15 years, there has been a notable increase in the allocation of supplementary grain-based concentrates (including grain/s, mixes of grains, commercial pellets) per cow [4–6]. Despite the potential for higher milk production (1 kg DM of concentrate contains 11-13 MJ of Metabolisable Energy (ME), which is equivalent to the ME required for 2 to 2.2 litres of milk), the average response of individual cows remains at approximately ~1 L of milk per kg DM of concentrate [7].

There are two major concentrate feeding strategies adopted by dairy farmers. The 'flat rate' feeding strategy, where every cow in the herd is allocated the same average amount of supplement whilst being milked, is used by about 78 % of dairy farms in Australia and New Zealand [8]. The alternative strategy is the individualised feeding system, where each

* Corresponding author.
E-mail address: blessing.azubuike@sydney.edu.au (B.N. Azubuike).

https://doi.org/10.1016/j.atech.2025.101669
Received 18 April 2025; Received in revised form 24 September 2025; Accepted 23 November 2025
Available online 24 November 2025

cow is fed differently and tailored to the nutritional requirements of individual cows. Only a few farms have adopted this strategy, despite previous research by García et al [9] reporting that adopting an individual cow feeding strategy ('feed to yield') increased milk production by up to 10 % under restricted forage conditions. This was achieved with the same amount of total concentrate used per day. However, this advantage was observed in the context of restricted forage intake, which differs from most dairy farmer who aim to feed forage ad libitum, explaining why many prefer flat-rate feeding strategies.

Individual supplementary feeding strategies may enable farmers to maximise milk productivity by tailoring supplementation to the specific nutritional requirements and responses of each cow, which vary both within and between cows [7,10,11]. Additionally, individualised feeding improves feed efficiency by allowing farmers to allocate more concentrate to high-yielding cows without overfeeding lower-producing ones, thereby optimising resource use. However, this approach must carefully manage the risk of negative cycles, where low-producing cows receive less concentrate and consequently yield even less. Even with the potential benefits and advancements in modern precision feeding technologies and automatic milking systems within pasture-based dairy operations (including computerised bail feeding systems and milk meters), farmers still lean towards the flat rate feeding strategy due to various challenges, as highlighted by Dela Rue and Eastwood [8]. Farmers find individualised feeding strategies complex, time-consuming, and capital-intensive, requiring automated feeders and silos [12]. While modern precision feeding technologies, including robotic milking systems, offer technical capabilities for automated yield-based allocation, implementation challenges related to management complexity, monitoring requirements, and economic considerations continue to influence farmer adoption decisions. In contrast, flat-rate feeding systems are easy to implement, save time, and provide equal supplement distribution, but they lack effective monitoring of individual cows feeding pattern, risking overfeeding or underfeeding and impacting pasture utilisation [11].

Retrospective analyses by André et al [13] have demonstrated economic potential for optimising concentrate feed based on individual cow variations in milk yield response, with benefits ranging from AU$0.33 to 3.34 per cow per day. However, the specific supplement rates and types responsible for these economic gains were not detailed in their study. A key limitation in much previous research [10,12–16] on feeding strategies has been the reliance on energy-requirement models, such as basic linear or logit models may be attributed to energy-requirement models. These models, while potentially accurate for average cow groups, often struggle to predict individual cow responses effectively, as they do not adequately capture the complex, non-linear relationships and increased variability among individual cows hence the need for approaches capable of more precisely modelling individual responses.

Data-driven approaches like machine learning (ML) can potentially overcome some of the above-mentioned limiting factors of the traditional 'biological-driven' approaches. Chauhan [17]; Liebe and White [15] emphasised that advanced non-parametric and non-linear models are likely to provide more accurate explanations for the variance in individual cow milk response [18–22]. These algorithms can integrate data from various sources at the individual cow level, revealing nonlinear patterns in cow behaviour within a herd. While machine learning models offer significant predictive benefits, optimisation challenges may require alternative approaches like mathematical programming or prescriptive analysis. Data-driven optimisation methods, proven effective in various domains like engineering, medicine, and finance, have the potential to enhance profitability and revenue by optimising resource allocation. However, their application in agriculture, particularly in dairy farming, remains relatively unexplored [23].

Mathematical optimisation, unlike traditional linear programming, could be a promising alternative as it uses advanced techniques for dynamic, precise supplement allocation, making it ideal for optimising concentrate feeding in pasture-based dairy systems [24]. We hypothesise that such a system can effectively utilise the allocated daily concentrate budget to predict optimal individual cow supplement allocations and maximise daily milk yield profitability [25,26]. This study addresses this gap by pioneering the integration of advanced machine learning techniques with mathematical optimisation to facilitate a dynamic, data-driven approach for individualised concentrate allocation in pasture-based dairy systems. Our approach uniquely leverages a highly performant non-linear model as a surrogate for milk yield prediction, for its ability to capture complex, non-linear relationships and inherent individual cow variability. This model is then embedded within a robust Differential Evolution optimisation framework, allowing for greater precision and computational efficiency compared to conventional methods.

Our study offers several unique insights and contributions. Firstly, we aim to investigate the quantifiable theoretical potential of this integrated approach to substantially increase herd-level milk yield without increasing the total concentrate budget, thereby enhancing profitability compared to conventional flat-rate methods. This directly addresses a key challenge in pasture-based systems where feed costs account for a significant portion (40-60 %) of milk production expenses. Secondly, our approach seeks to achieve significant enhancement in predictive accuracy by integrating both concentrate levels and individual cow physiological characteristics (days in milk, liveweight, and parity number). This capability is vital, as it allows the model to move beyond group-level analyses and capture the non-linear responses of individual cows, recognising that the relationship between concentrate and milk yield exhibits diminishing returns as concentrate levels increase. Finally, this research strives to provide a foundational framework for tactical decision support systems in modern dairy farming, offering a scalable and less time-consuming alternative to manual or traditional rule-based feeding strategies. It highlights how automated feeding systems can benefit from real-time data-driven adjustments to optimise feed utilisation and promote animal welfare.

2. Materials and methods

The main objective of this study is to maximise herd-level daily milk yield by optimising the daily allocation of concentrate feed among a fixed set of cows, while maintaining approximately the same total daily feed quantity. To achieve this, a two-step methodology was proposed. The first step involved development a machine learning model that uses cow characteristics such as days in milk (DIM), liveweight (LW), number of lactations (NOL) and concentrate consumption or concentrate level (CONC) to predict individual milk yield (MY). The second step involved optimising the randomly simulated initial concentrate levels using the developed predictive milk yield model.

2.1. Data

Data were collected from a controlled field study conducted with 130 lactating Holstein-Friesian primiparous and multiparous cows at the Corstorphine Dairy Farm Camden campus, University of Sydney, Australia between October 2010, and January 2012. The cows were enrolled as they calved with 42 cows enrolled during spring 2010, 15 cows during summer 2010, 17 cows during winter 2010, 14 cows during autumn 2011, and 42 cows during summer 2011. All the cows were managed as a herd and were milked twice daily in a 20-aside herringbone dairy. At every milking, the cows were fed commercial pellets of cereal-based concentrate formulated to 18 % CP containing 600 g/tonne available zinc (60 ppm in feed + 170 ppm endogenous in the premix = 230 ppm), 200 g/tonne Biotin 2 % (4 ppm in feed) and 420 g/tonne Rumensin 100 (42 ppm monensin in feed). The concentrate had a metabolisable energy content of 13.1 MJ/kg DM, with 16.8 % neutral detergent fiber (NDF), 8.1 % acid detergent fiber (ADF), 10.1 % Ash, 3.7 % N and 5.3 % water-soluble carbohydrates. The concentrate levels allocated to the individual cows ranged between 8 and 10 kg/cow/day

which was based on individual cow milk production level (<30 L/d: 8 kg; 30 to 35 L/d: 9 kg; >35 L/d: 10 kg). This allocation method reflected management decisions linked to existing production levels rather than an experimental manipulation of concentrate allowance and thus describe correlations under standard practice rather than direct cause and effect responses to varied concentrate levels. Individual animal milk yields and liveweight were automatically recorded at each milking using milk meters and automatic scales respectively, both tested for accuracy and calibrated weekly.

The raw data collected for each of the 130 cows were organised into separate .csv files provided by the Delpro software (DeLaval, Sweden), with each file corresponding to daily records of individual milk yield, reproduction, and feeding. After merging these files, a data cleaning process was applied, which included the removal of outliers, records with missing data, imbalances, and instances with zero milk yield entries. This filtering resulted in a refined dataset consisting of 32,504 observations, systematically recorded over a continuous period of 492 days, from September 3, 2010, to January 22, 2012. To facilitate a comprehensive examination of the distribution and enrolment of cows throughout the trial, timestamps indicating sampling dates and the "days in milk" values were discretised and categorised into distinct seasons of the year. The stages of lactation were defined as early (1 to 100 days in milk), mid (101 to 199 days in milk), and late lactation (200 or more days in milk). From the curated dataset DIM, stage of lactations (SOL), NOL, LW, CONC, MY and Cow Identification Number (Cow ID) were selected for the study.

All data pre-processing and analyses were conducted using the open-source statistical software package R, version 4.3.1.

2.2. Data analysis

A correlation analysis was performed to explore relationships between the variables. Spearman's method, priced for its non-parametric approach and focus on monotonic relationships rather than linear ones, was employed for this purpose. The Shapiro-Wilk test was utilised to assess the normality of the distribution of the variables in the dataset and to evaluate the non-linear connections between the variables. This preliminary analysis provided essential insights into the distribution of each variable and their interrelationships, informing the subsequent comprehensive modelling approach.

In the preprocessing phase, the dataset was divided into numerical / continuous (DIM, LW, CONC) and non-numerical / categorical (NOL, SOL and Cow ID) features. The categorical variable "NOL" underwent One Hot Encoding, transforming it into an integer format suitable for diverse modelling algorithms. For the numerical features, to address issues related to feature scale and distribution, standardisation was used to centre the features on a mean of 0 with a standard deviation of 1, balancing their impact, improving training efficiency, and preventing scale-related bias during the learning process.

2.3. Step 1: Predictive model for milk yield

The development, evaluation and optimisation of the predictive milk yield model were facilitated by using the Python (v3.11.4) programming language and its versatile libraries. The predictive model developed in this study included MY as the target variable for prediction and DIM, LW, NOL and CONC as predictor variables. To ensure robust validation, the dataset was randomly split into training and test subsets at the individual cow level, with 80 % of records for each cow used for training and 20 % for testing. To account for randomness in the splitting process, multiple random iterations of the train-test split were performed using a fixed random state of 42 to ensure reproducibility. The split was balanced across key variables including days in milk (DIM), parity, and milk production to ensure both training and test sets reflected the diversity of the dataset. Additionally, since cows were enrolled at different time points, the split maintained temporal representativeness by preserving the chronological structure of records within each cow. This resulted in 26,003 observations used for training and 6501 observations used for testing. The training set was utilised to fit the model and determine optimal weights and coefficients, while the test set was kept for an unbiased evaluation of model performance.

To evaluate and underscore the importance of individual cow characteristics and feed in refining the precision and accuracy of MY predictions, three distinct modelling procedures were implemented. In the first procedure the predictive factor solely relied on CONC. The second procedure excluded CONC and focused solely on individual cow characteristics. Subsequently, the third procedure integrated both individual cow characteristics and CONC to estimate daily individual cow MY comprehensively.

2.3.1. Model selection

To determine the most effective machine learning algorithm for predicting MY, the performance of sixteen ML techniques comprising of seven linear and nine non-linear algorithms (Table 1) was compared. The evaluation metrics for assessing model accuracy included Mean Absolute Error (MAE), Mean Squared Error (MSE) or Root Mean Square Error (RMSE), R Squared (R^2) and Mean Absolute Percentage Error (MAPE). Random Forest (RF), the algorithm demonstrating the highest R^2 value and the lowest RMSE, MSE and MAE, was selected as the best model for MY prediction. The R^2 value indicates the proportion of variances explained by the model, thereby reflecting its accuracy and the relevance of cattle characteristics in MY modelling.

2.3.1.1. Random forest algorithm. Random Forest is an ensemble machine-learning algorithm that employs decision trees for both regression and classification tasks. It combines the predictions of multiple micro-scale decision trees, also known as estimators, using bagging or bootstrap aggregation [47–49]. The RF model was selected for comparison due to its ability to mitigate overfitting and enhance the generalisation of predictions.

To systematically explore the parameter space for the optimal RF model configuration, a grid search approach combined with a k-fold cross-validation technique (with k set to 5) was implemented. This comprehensive search evaluated nine different numbers of estimators, ranging from 100 to 500 with increments of 50. The Negative Mean Squared Error (NMSE) served as the evaluation metric during the grid search process to gauge predictive performance and facilitate model selection.

Following the grid search, the RF model was fitted using the number of estimators associated with the highest NMSE value. This optimised model was then applied to the test dataset for performance evaluation. Furthermore, to discern the variables exerting the most significant influence on MY predictions, feature importance analysis was conducted. This technique provided a quantitative measure of the contribution of each feature to the predictive power of the model, thereby enabling a deeper understanding of the factors driving the output of the model. Through this analysis, insight was gained into the relative importance of each variable within the RF model framework.

2.4. Step 2: optimisation of individual concentrate levels

To optimise individual concentrate levels, a Monte Carlo simulation was integrated with the Dirichlet-Rescale (DRS) algorithm and milk yield predictions to establish initial values for daily individual feed allocations while respecting all set constraints. Monte Carlo simulation generated a wide range of possible scenarios for feed allocation, enabling exploration of various ways to distribute the feed among the cows. The Random Forest-based milk yield model, used as a surrogate model, estimated milk yield for the feed allocations generated by the Monte Carlo simulation. The DRS algorithm adjusted those allocations, ensuring that the total daily feed quantity remained the same and that

Table 1
Summary of all the machine learning models assessed in the study.

Categories	Group Type	Model Used for the Study	Optimisers Used	Literature
Linear Algorithms (Parametric models)		Multiple Linear Regression (LR)		[27–30]
		Least Absolute Shrinkage and Selection Operator (LASSO)		[31]
		Ridge Regression (RR)		[32]
		Elastic net Regression		[33]
		Polynomial Regression		[34,35]
		Partial Least Squares Regression (PLS)		[36,37]
		Support Vector Machine (SVM)		[38,39]
Non-linear Algorithms (non-parametric models)	Gradient Boosting Algorithms	Gradient Boosting Machines (GBM)		[40,41]
		eXtreme Gradient Boosting algorithm (XGBoost)		[42]
		Light Gradient Boosting Machine (LightGBM)		[43]
		Categorical Boosting (CatBoost) algorithm		[44]
	Tree-based machine learning algorithms	Decision Tree ensembles		[45,46]
		Random Forest (RF) ensembles		[47–49]
	Other non-parametric models	K-Nearest Neighbours algorithm (KNN)		[50,51]
		Gaussian Process (GP)	Adaptive Moment Estimation (Adam)	[52–56]
	Recurrent Neural Network (RNN)	Artificial Neural Networks (ANN)	Root Mean Squared Propagation (RMSProp)	[57]
			Stochastic Gradient Descent (SGD)	[58,59]
			Adaptive Moment Estimation (Adam)	[60]

each cow received an appropriate amount of feed based on their characteristics. Subsequently, these selected initial values were leveraged alongside the SciPy open-source Python library to compute the optimal individual daily feed allocations. An illustrative representation of this feed optimisation process is provided in Fig. 1.

The employment of the pre-trained RF model as a surrogate (ML) model was undertaken to reduce computational complexity, drawing on previous research by Franzoi et al [61]; Wang et al [62]; Williams and Cremaschi [63]. Surrogate models are often used to approximate complex systems where direct simulations or computations are expensive, and map input data to outputs via black-box modelling. This approach allowed us to efficiently integrate cow-specific traits and optimise MY predictions without the need for computationally intensive simulations.

The optimisation of individual concentrate levels was done on a subset of data extracted from the pre-processed dataset, comprising 7371 records of 81 cows collected over 91 consecutive days, from April 5, 2011, to July 8, 2011. The subset included 27 first-lactation, 17 s-lactation, 19 third-lactation, and 18 fourth-lactation cows, categorised by NOL of the cows. Additionally, the cows were differentiated based on their stages of lactation, with 42 cows in early lactation, 13 in mid-lactation, and 26 in late lactation. This was done to ensure a minimum period for which a subset of cows (and their records) would be available, despite cows having calved at different times.

2.4.1. *Monte carlo simulation for initial concentrate allocation*

For each of the 81 cows, random quantities of concentrate dry matter representing simulated daily concentrate levels, were generated through the deployment of a Monte Carlo simulation approach. This entailed executing 500 iterations per cow per day to ensure a comprehensive exploration of potential feed allocation scenarios. Then, the DRS algorithm was applied to ensure that the generated concentrate levels conformed to two critical conditions. First, the sum of the generated concentrate levels for each iteration equalled the total actual concentrate allocation which was derived from the observed concentrate levels allocated to the cows from the dataset. Second, each concentrate level for a cow was constrained within the specified limits of the minimum and maximum allowable concentrate levels (5 kg/cow.day^{-1} and 9 kg/cow.day^{-1}, respectively). Following this, a sensitivity analysis was conducted to assess the impact of varying concentrate level constraints on MY and the optimal allocation strategy.

Across the 500 iterations, predictions of MY for each cow were computed using a pre-trained RF model, leveraging the corresponding simulated concentrate levels generated for each scenario. For each iteration, the total MY was ascertained by summing the individual predictions for all 81 cows. This process was designed to evaluate the effects of various concentrate feed allocations on the collective milk production output, thereby enabling a nuanced understanding of how feed adjustments could potentially optimise overall dairy yields. Ultimately, the set of simulated concentrate levels that produced the highest total MY was pinpointed and selected as the initial reference point for the optimisation process, conducted within the confines of the predefined constraints (Fig. 1).

2.4.2. *Optimisation of initial concentrate for milk yield maximisation*

The optimisation problem was addressed using the “minimise” function from the SciPy library, employing the “differential evolution” method, which handles both inequality and equality constraints. Differential Evolution (DE), a type of evolutionary algorithm, was chosen for its ability to avoid getting trapped in local optima through its stochastic approach. Unlike conventional gradient-based techniques e.g. linear programming, DE algorithm can efficiently explore a wide candidate space with numerous function evaluations, making it suitable for complex optimisation problems. An objective function was formulated to maximise the sum of the predicted milk yields obtained with the decision variables using the best-trained machine learning model. The decision variables represented the simulated concentrate levels for each cow and contained 81 values. The objective aimed to maximise the individual cow MY predictions simultaneously. Constraints were imposed to ensure that the sum of the simulated concentrate levels must be at least 2 % less than the total actual concentrate or equal to the total

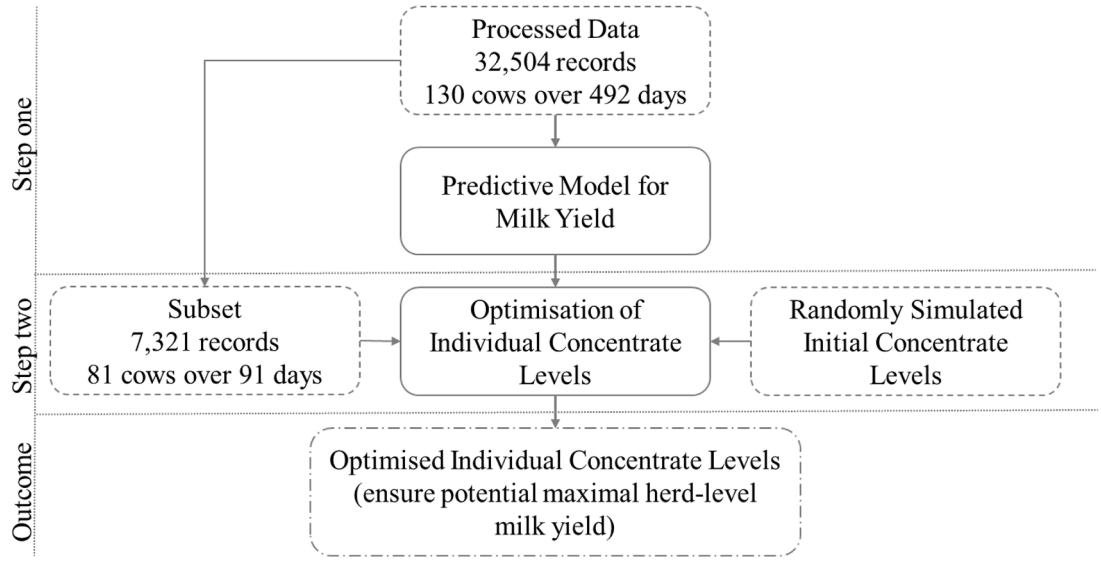


Fig. 1. Illustration of the two-step feed optimisation process: data is represented by dashed rectangles, models by solid rectangles, and outputs by dashed-dotted rectangles.

concentrate allocation observed in the dataset for each day. Additionally, constraints were applied element-wise to enforce that each individual allocated concentrate by the model should not exceed >1 kg/cow.day^{-1} from one day to the next day for each cow. Bounds on concentrate allocated per cow was also reinforced. The initial guess for the decision variables was set to the initial concentrate values selected from the Monte-Carlo simulations. A callback function was defined to store the predictions and concentrate levels for each iteration for evaluation purposes (Fig. 1).

2.4.3. Equation for the optimisation problem

Mathematically, the optimisation problem can be represented as a maximisation of the following objective function and constrains:

$$\text{Maximise}: \quad -\sum_{i=1}^{n} y_i$$

Subject to:

$\text{Total\_Concentrate\_Allocation} * 0.98 \leq \sum_{i=1}^{n} x_i \leq \text{Total\_Concentrate\_Allocation}$

$y_i \leq \text{Actual Milk Yield}_i \times 1.1$, for cow i

$5 \leq x_i \leq 9$, for each cow i

Where: n is the number of cows in the dataset. x_i is the simulated concentrate level for cow i. y_i is the predicted MY for cow_i obtained from the machine learning model.

Total_Concentrate_Allocation is the total amount of concentrate available for all cows.

Actual Milk Yield is the actual MY for cow_i obtained from the dataset.

The optimisation process aimed to find the values of x_i (simulated concentrate levels) that maximise the total MY $\sum_{i=1}^{n} y_i$ while respecting the individual maximum increment constraint and the sum constraint on the simulated concentrate levels, as well as the bounds on the simulated concentrate levels.

In addition to setting the total concentrate level constraint between 98 % and 100 % of the initial allocation, other constraints were also tested, including 90 %−100 %. These variations were implemented to assess the optimisation sensitivity of the model to different tolerance ranges and to understand how these constraints affect computation time and complexity. By exploring these different ranges, the goal was to determine how the constraints influence the time required to find the optimal solution and the overall performance of the model under varying conditions.

3. Results

3.1. Data analysis

3.1.1. Descriptive analysis

The dataset captured a wide range of individual cow records, varying from 23 to 289 observations per cow, with only twelve out of the 130 cows having fewer than 200 records. While it provided extensive coverage for DIM, LW, and MY, the allocated concentrate levels exhibited less variability, with values ranging from 8 to 10 kg/cow.day^{-1} and an average of 8.27±0.57 kg/cow.day^{-1} (Table 2). This modest variability in the allocated concentrate levels may be considered as a potential limitation of the dataset. Despite this, the MY per unit of concentrate varied broadly, with a range from 0.97 to 6.93 L/kg of concentrate/cow.day^{-1}, and an average of 3.38 ± 0.70 L/kg of concentrate/cow.day^{-1}, indicating the potential for optimising concentrate use to maximise MY. This variability in milk yield per kg of concentrate in this dataset is primarily explained by the individual cows underlying milk production level, influenced by factors such as genetic potential and physiological stage, rather than directly reflecting the utilisation efficiency of concentrate.

Examining the distribution of cows by lactation stage, approximately 27 % were in early lactation (<100 DIM), while a similar proportion (~36 %) fell into mid and late lactation stages (>100 DIM). With respect to the NOL, 34.9 % were in their first, 20.8 % in their second, 21.6 % in their third, and 22.8 % in their fourth or subsequent lactation.

Notable influences of lactation stage and the NOL on milk production were observed. Cows in early lactation had a higher average MY of 31.48 ±6.79 L/cow.day^{-1}, compared to 27.81±5.97 L/cow.day^{-1} for mid-lactation and 25.65±5.75 L/cow.day-1 for late lactation (Appendix

Table 2
Summary statistics of the days in milk (DIM), lightweight (LW), concentrate consumed (CONC) and milk yield (MY).

Metrics	DIM (days)	LW (kg)	CONC (kg/cow)	MY (L/cow)
Minimum	7.00	312.00	8.00	8.00
25 %	95.00	529.00	8.00	23.58
Mean	162.55	586.73	8.27	28.03
Median	162.00	589.00	8.00	27.70
75 %	231.00	643.00	8.00	32.10
Maximum	310.00	886.00	10.00	55.40
Std Dev	80.09	75.11	0.57	6.54

Table A 1). First and second lactation cows produced approximately 79 % and 95 %, respectively, of the MY of cows in their third and fourth lactations (~30.6 L/cow.day-1).

3.1.2. Correlation and distribution assessment

Correlation analysis revealed that among the predictor variables for MY, the concentrate level and the NOL showed the strongest positive correlation (0.44) with MY (Fig. 2). Conversely, and as anticipated, DIM was inversely (−0.37) related to MY. Notably, LW also showed a positive correlation (0.31) with MY, albeit to a lesser extent than concentrate levels and lactation history.

Furthermore, the Shapiro-Wilk test revealed that the input variables (NOL, DIM, LW, CONC) deviate significantly from a normal distribution (p-values substantially below 0.05), confirming their non-Gaussian distribution patterns. Notably, LW exhibited a pronounced deviation from a normal distribution.

3.2. Predictive model for milk yield

3.2.1. Performance of machine learning algorithms

Among the sixteen machine learning algorithms (Table 1) evaluated for predicting MY based on NOL, DIM, LW, and CONC, the RF model demonstrated superior performance yielding an R^2 of 0.6 and RMSE of 4.2 L/cow.day^{-1} (Table 3). Conversely, the Robust Regression algorithm exhibited the lowest predictive accuracy, attaining an R^2 of 0.30 and RMSE of 5.60 L/cow.day^{-1}. Generalised Linear Models (GLMs) such as linear regression, ridge regression, lasso regression, elastic net regression, and robust regression exhibited relatively lower predictive accuracy when compared to non-linear models such as Gradient Boosting Machines (GBMs), Gaussian Process Models, RF ensemble methods, and neural networks which demonstrated superior performance (Table 3). These results highlight the benefits of using non-linear models like RF to identify these patterns over traditional GLMs, enabling more precise modelling procedures and decision-making in dairy production.

3.2.2. Hyperparameter optimisation

A thorough grid search assessment and hyperparameter optimisation were conducted to select the most suitable predictive model. The optimal hyperparameters for training the final RF Regressor model on the designated training datasets were determined to be 350 n-estimators [64]. These parameters were chosen based on their ability to provide the best balance between accuracy and computational efficiency, resulting in an average minimum MSE of 17.68, alongside an RMSE of 4.204 as illustrated in Fig. 3(a, b). The training dataset was then utilised to fit the RF model with the optimised parameters, serving as the foundation for evaluation and further analysis

Table 3
Performance comparison: machine learning models used for milk yield prediction.

Machine Learning Algorithms	R^2	MSE	RMSE (L/cow. day^{-1})	MAE	RMSLE
Linear Regression	0.450	24.345	4.934	3.803	0.189
Ridge Regression	0.450	24.347	4.934	3.803	0.189
Lasso Regression	0.370	27.904	5.282	4.088	0.201
Elastic Net Regression	0.447	24.507	4.951	3.826	0.190
LightGBM	0.502	22.038	4.695	3.623	0.181
PLS Regression	0.450	24.345	4.934	3.803	0.189
Robust Regression	0.301	30.945	5.563	4.325	0.202
Polynomial Regression	0.499	22.191	4.711	3.630	0.182
Gradient Boosting Regression	0.584	18.440	4.294	3.241	0.167
XGBoost	0.504	21.990	4.689	3.626	0.181
Decision Tree	0.371	27.877	5.280	3.714	0.207
KNN Regression	0.433	25.117	5.012	3.886	0.187
Support Vector Machines (SVM)	0.539	20.407	4.517	3.411	0.175
Random Forest Regressor (RF)	0.602	17.625	4.198	3.061	0.164
Neural Network (Adam)	0.583	18.494	4.300	3.312	0.164
Neural Network (SGD)	0.537	20.502	4.528	3.488	0.174
Gaussian Process (SVGP)	0.597	17.862	4.226	3.216	0.164
Gaussian Process (Exact GPs)	0.548	20.032	4.476	3.441	0.173

Notes: R^2 = *R*-squared, MSE = Mean Squared Error, RMSE = Root Mean Squared Error, MAE = Mean Absolute Error, RMSLE = Root Mean Squared Log Error.

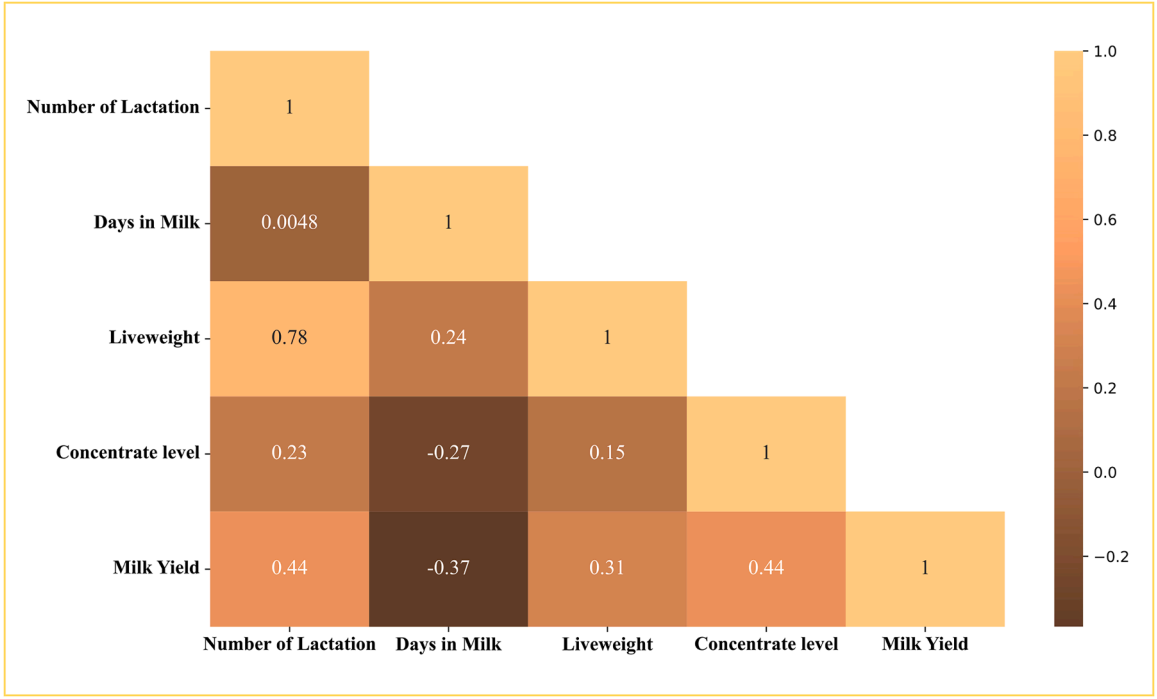


Fig. 2. A correlation heatmap illustrating the relationships between the variables in the dataset.

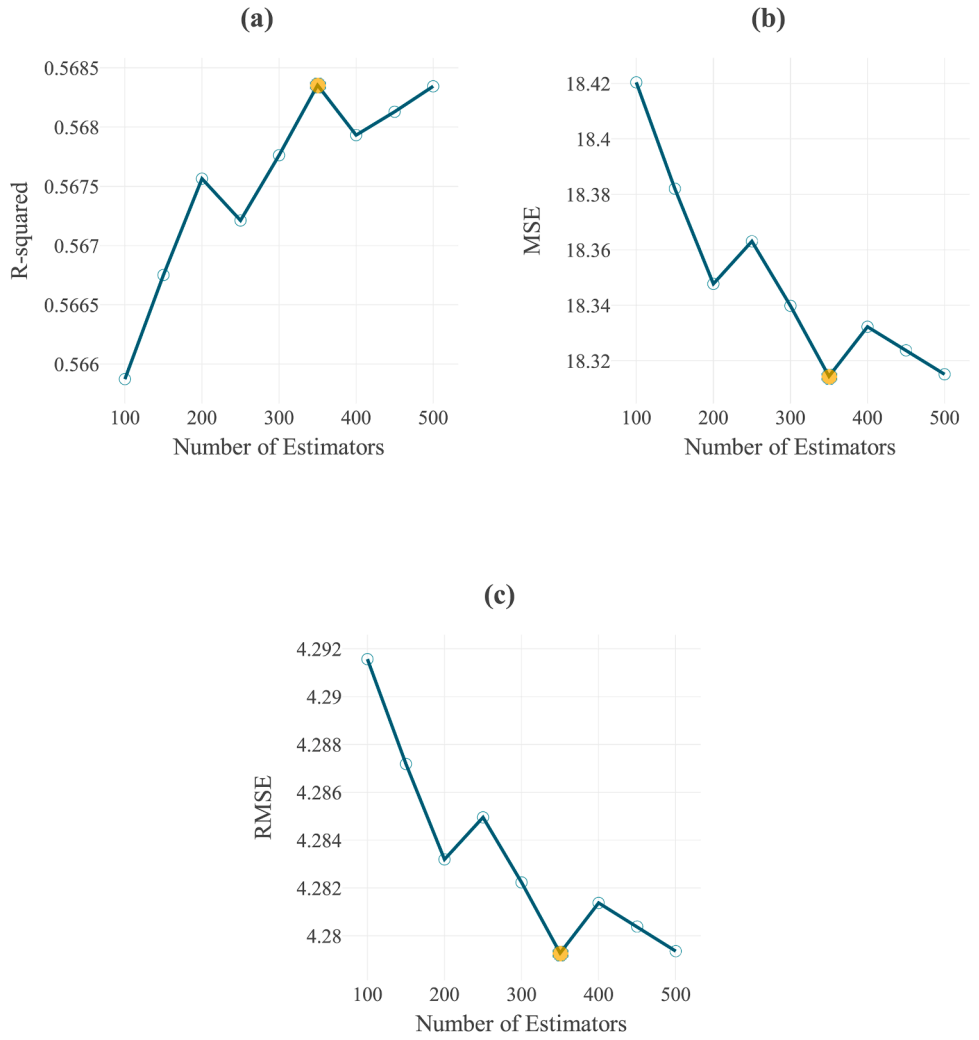


Fig. 3. Impact of the number of estimators on the performance of the RF model': R-squared (a), Mean Squared Error (b) and Root Mean Squared Error (c) Orange shaded markers highlight the optimal number of estimators (n_estimators = 350) used for fitting the training dataset during milk yield prediction.

3.2.3. *Predictive model efficacy*

Fig. 4 presents the critical role of intrinsic cow characteristics alongside feed data in the prediction of individual cow milk yields. When considering only the grain-based concentrate (Fig. 4a), the predictive power of the model was very low, as indicated by an R^2 value of 0.29. Focusing solely on the characteristics of the cows (Fig. 4b) provided a marginally improved fit, with an R^2 value of 0.31. However, integrating both the concentrate levels and the individual cow characteristics into a composite predictive model (Fig. 4c) significantly enhanced the explanatory power of the model, achieving an R^2 value of 0.60. This notable improvement in predictive accuracy highlights the complex relationship between the distinct attributes of individual cows and their dietary intake, which collectively shape the daily MY outputs.

The results also show the substantial impact of constructing individualised predictive models for MY at the cow level. Notably, the RF predictive efficacy of the model varied significantly based on the parity of the cows. Primiparous cows (NOL = 1) demonstrated the lowest predictive accuracy, with approximately 26 % exhibiting R^2 values ranging from 0 to 0.25 while predictive accuracy increased substantially for multiparous cows.

3.3. *Grain-based feed optimisation*

In the dataset extracted for feed optimisation purpose, a total daily concentrate consumption for 81 cows ranged from 651 to 732 kg.day^{-1}, and total daily milk production ranged from 1881 to 2483.6 L.day^{-1}. For the entire period of 91 days, 60,625 kg of concentrate was used, resulting in 199,488.20 L of milk and the efficiency of milk production varied from 1.15 to 5.31 L/kg of concentrate/cow.day^{-1}, with an average of 3.29 L/kg of concentrate/cow.day^{-1} (Fig. 5, green line). The optimisation process resulted in an average improvement of 8.11 % in total daily MY, with daily increments ranging from 2 % to 23.4 % when the constraints for total optimised concentrate was set to be between 98 % and 100 % of the total actual concentrate daily (Fig. 5, orange line). The results indicate that periods of substantial declines in actual MY corresponded to proportionately larger percentage increases in MY during subsequent optimisations.

A detailed analysis of milk production trends (Fig. 5) shows distinct patterns in herd MY corresponding to changes in concentrate allocation. Notably, the lowest herd milk production levels were observed around mid-May 2011, followed by a marked increase when the total daily concentrate was raised by approximately 80 kg, reaching 732 kg. This increase, which adjusted individual cow allocations from 8 kg/day to 9 kg/day, resulted in a significant rise in the predicted total MY, peaking at 2528.5 kg/day (Fig. 5). On average, the model predicted a total MY of 2367.03 ± 84.48 kg/day, which is higher than the actual recorded average of 2192.18 ± 115.17 kg/day over the 91-day period.

Fig. 6 illustrates the optimisation effects when the constraint on the total optimised concentrate was set to remain between 90 % and 100 % of the initial allocation. This constraint ensured a steady flow in the

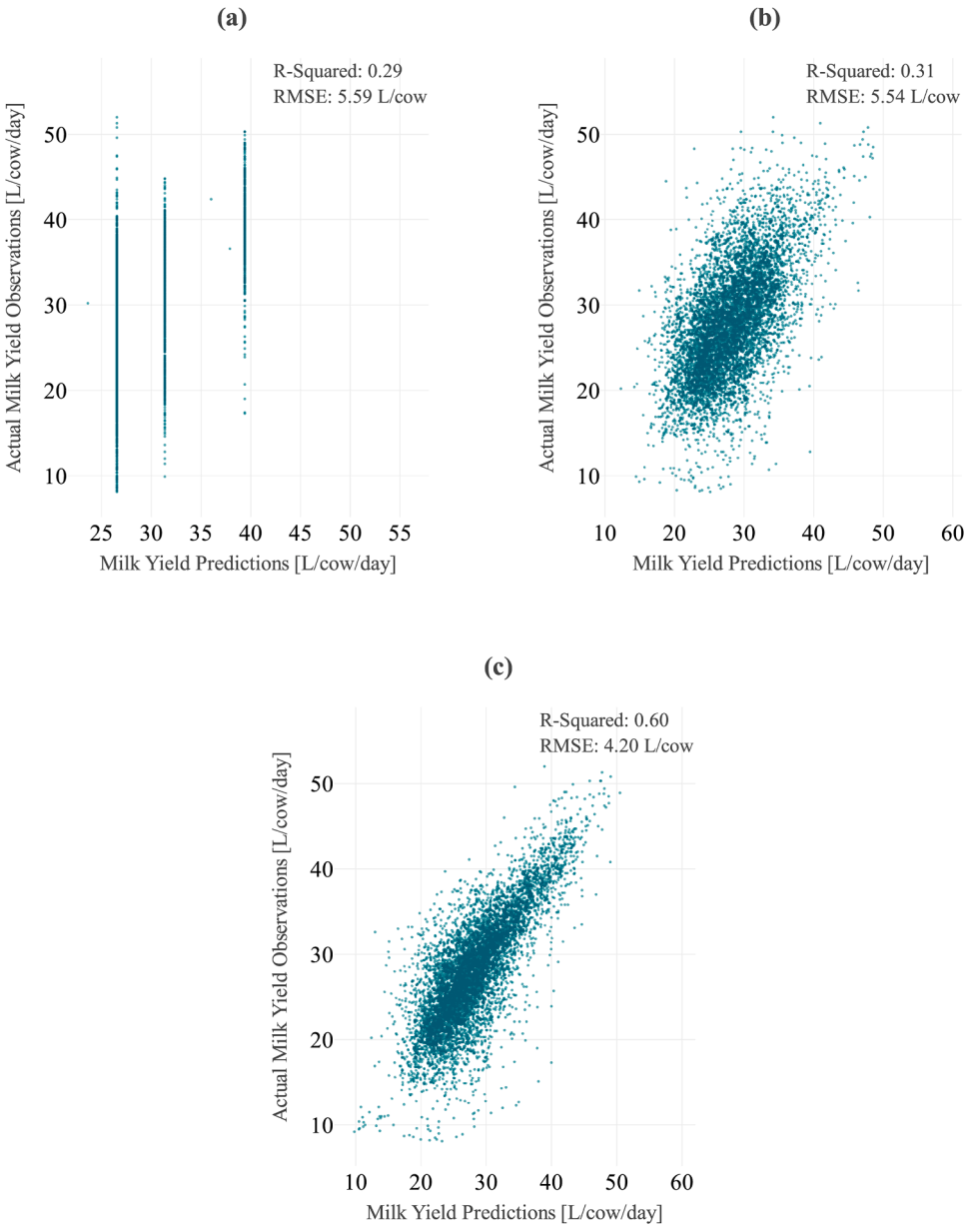


Fig. 4. The importance of feature variables for individual milk yield modelling: association between predicted and actual milk yield when input to model is a) only concentrate; b) only cow characteristics; and c) all predictors: cow characteristics and concentrate. Here, MSE = Mean Squared Error and RMSE = Root Mean Squared Error.

optimisation process, maintaining consistency across the sampling period. The analysis revealed that the percentage increase in MY ranged from 0.2 % to 15.5 % per day over the 91-day period. Notably, there was no significant spike in the total maximised MY on any particular day, including mid-May 2011. The average percentage increase in MY was 6.34 % across the 91 days, demonstrating the effectiveness of the model in enhancing milk production within the specified constraints. These results show the importance of maintaining a controlled range for concentrate allocation, as it leads to more predictable and stable improvements in MY. The consistent flow observed in the optimisation process further highlights the reliability of the model. Overall, the results from Fig. 6 confirm that setting appropriate constraints on concentrate levels can optimise milk production without causing abrupt fluctuations. This approach ensures a sustainable increase in MY, benefiting both the health and overall dairy production efficiency of the herd.

When compared to Fig. 5 which depicts the optimisation results without specific constraints on concentrate levels, it is evident that constraining the concentrate allocation (as shown in Fig. 6) led to a more consistent and predictable optimisation process. While Fig. 5 showed average improvements of 8.11 % with daily increments ranging from 2 % to 23.4 %, Fig. 6 indicated a slightly lower but more stable average increase of 6.34 %. Overall, the comparison between Fig. 5 and Fig. 6 confirms that while more relaxed constraints can lead to higher maximum increases, maintaining tighter constraints provides more predictable and steady improvements in MY.

From our model, it was observed that this approach suggests substantial increases in daily MY when cows are fed individually, compared to those under static flat rate allocations indicates the potential of optimising milk production by adjusting feed concentrations to align precisely with the unique nutritional needs and responses of each cow.

To provide a more granular understanding of the optimisation process, Fig. 7 presents a breakdown of the daily sample results, detailing the day-to-day adjustments for each cow. This breakdown illustrates

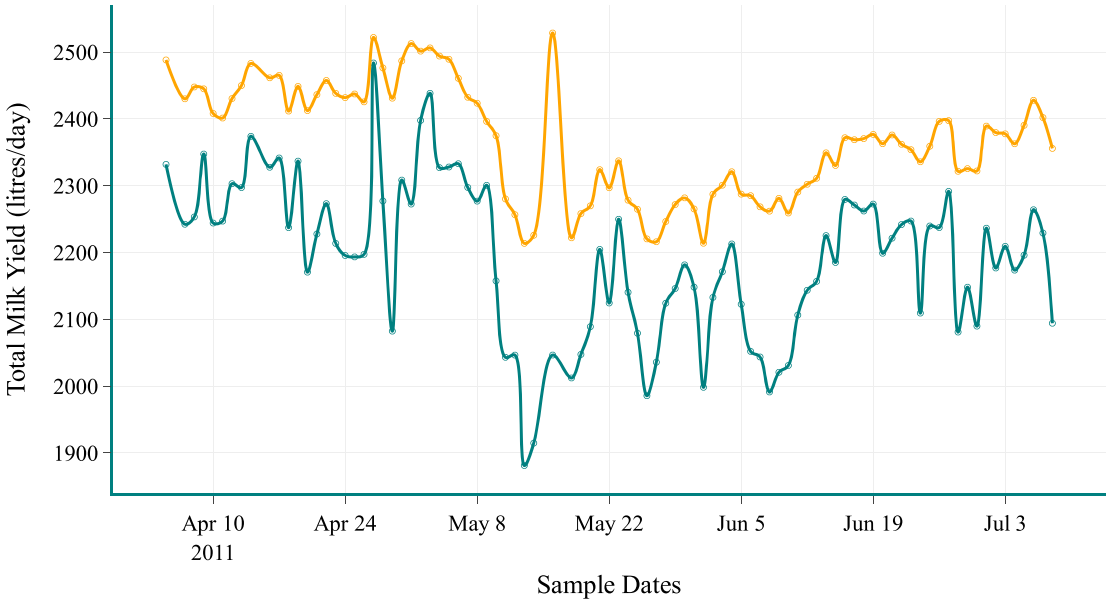


Fig. 5. Total daily milk yield of 81 cows before (green line) and after (orange line) optimisation across 91 days. All constraints applied at 98-100 % range for total actual concentrate consumed. The annotation on the orange line shows the daily achieved milk yield increase by the optimisation model.

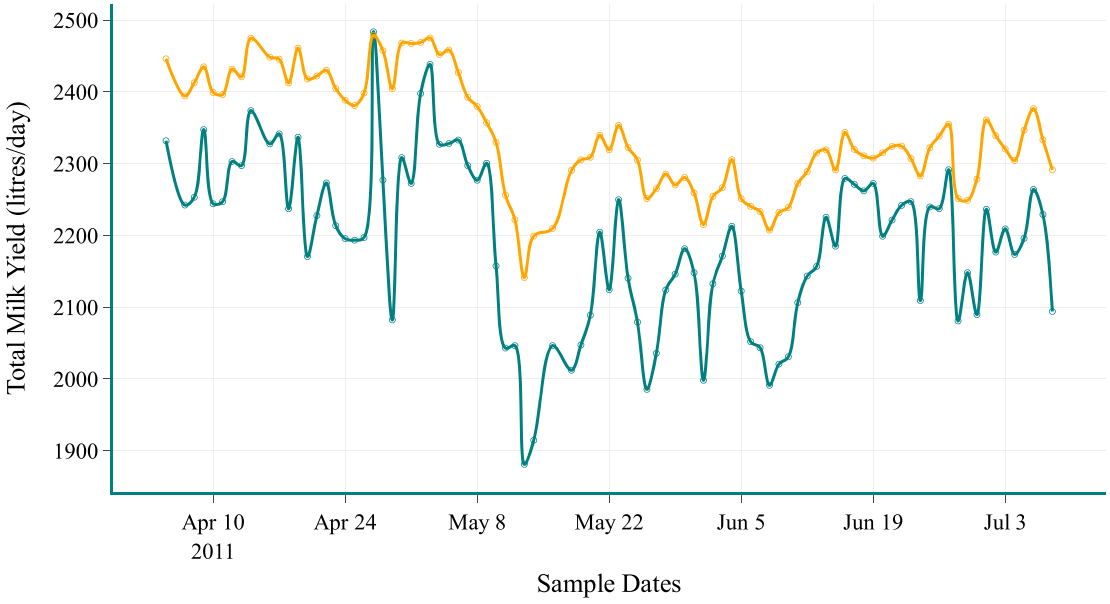


Fig. 6. Total daily milk yield of 81 cows before (green line) and after (orange line) optimisation across 91 days. All constraints applied at 90-100 % range for total actual concentrate consumed. The annotation on the orange line shows the daily achieved milk yield increase by the optimisation model.

how the model dynamically responds daily to varying conditions, optimising concentrate allocation and MY predictions. It highlights the adaptability and precision of the model in achieving optimal outcomes for each individual cow, providing a closer examination of its effectiveness in practical, cow-level scenarios.

3.4. Correlation analysis of concentrate allocation adjustments

To understand how the optimisation algorithm redistributed concentrate among cows with different characteristics, a correlation analysis was conducted between the change in concentrate allocation (Delta_Concentrate = Optimal - Actual) and individual cow traits. The analysis revealed that concentrate allocation adjustments were not strongly associated with individual cow characteristics. DIM demonstrated a weak but statistically significant negative correlation ($r = -0.036$, $p = 0.002$), whilst LW ($r = 0.009$, $p = 0.431$) and NOL ($r = 0.001$, $p = 0.923$) showed no significant correlations with allocation changes. The Delta concentrate values averaged -0.290 ± 0.967 kg/cow/day, indicating that the optimisation algorithm typically reduced individual cow allocations whilst redistributing concentrate more efficiently across the herd, with changes ranging from -4.618 to $+1.000$ kg/cow/day. Significant group differences were observed across SOL ($F = 10.367$, $p < 0.001$) and parity ($F = 4.393$, $p = 0.004$). Cows in mid-lactation experienced the smallest average reduction (-0.237 kg/cow/day) compared to early lactation (-0.354 kg/cow/day) and late lactation (-0.317 kg/cow/day). Analysis by parity revealed that second lactation cows had the smallest reduction (-0.227 kg/cow/day), whilst third lactation cows experienced the greatest average reduction (-0.344

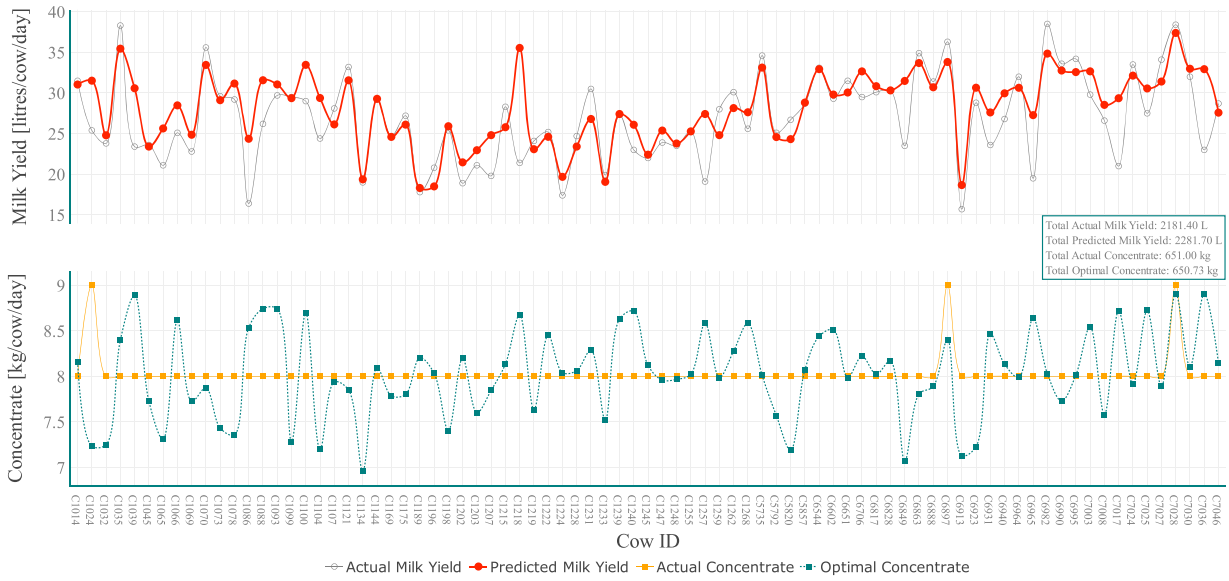


Fig. 7. Optimisation results for one of the sample dates (2011-05-30) used in the study. All constraints applied at 98-100 % range for total actual concentrate consumed. The orange line shows the actual concentrate level for each cow for the day, the green line shows the optimised concentrate for each cow for the day. The gray line shows the actual milk yield for each cow for the day and the red line shows the achieved milk yield for each cow per day by the optimisation model.

kg/cow/day).

4. Discussion

The aim of our study was to develop an optimisation model capable of individually distributing grain-based concentrate to cows to optimise milk production without increasing total concentrate used per day. To improve our model development and reduce the computational time associated with conventional optimisation models, we integrated a non-linear ML model into the optimisation process. This approach allowed for more accurate predictions of MY on a cow-by-cow basis, leveraging data-driven insights to fine-tune concentrate allocations. Our results indicate that this integration led to an approximate 6 to 8 % increase in total MY at the herd level.

4.1. Milk yield predictive modelling

A comparative analysis of diverse machine learning modelling techniques, as examined in prior research [65–67], reveals that non-linear models exhibit superior performance over regularised linear models when it comes to explaining the variance in MY for individual cows. The performance evaluation of ML algorithms for MY prediction in this study showed similar findings. Particularly, all the GLMs exhibited relatively lower predictive accuracy when compared to non-linear models such as the selected RF model.

These results affirm the limitations reported by Liebe and White [15] regarding the challenges linear models face in managing complex, non-linear relationships within data. A strictly linear relationship between MY and concentrate allocation would suggest that more concentrate always increases milk production, but in reality, this relationship is non-linear. As concentrate levels rise, diminishing returns occur, where increases in concentrate do not yield proportional gains in milk production. However, non-linear machine learning models like RF offer greater flexibility and robustness in inherently capturing these complex patterns.

Our study highlights the importance of individual cow characteristics and feeding data in predicting MY, corroborating previous research that also affirms the potential to explore non-linear relationships within and between cow variations [22,68]. There were also insights on the benefits of exploring the impact of individual cow-level variability in MY prediction. The analysis showed constructing individualised predictive models for MY at the cow level and emphasises the sensitivity of the predictive model to individual cow characteristics, implying that the model adjusts its predictions based on the distinct variability observed within and between cows.

While earlier studies [22,69] have incorporated variables such as starch content, crude fibre, PDIE, parity, and the number of milkings per day for MY prediction using machine learning algorithms like SVM, RF, GBM and ANN, however, our investigation specifically focused on four predictor variables: DIM, LW, NOL and concentrate level. Our results show the importance of integrating cow-specific traits to enhance predictive accuracy of the model. The integration of physiological traits into the MY prediction model resulted in a reduction of RMSE from 6L/cow.day^{-1} to 4L/cow.day^{-1} (Fig. 4 and 4b).

This reduction in error aligns with Salamone et al [70], who also demonstrated improved accuracy when incorporating cow-specific factors resulted in a better model prediction performance with RMSE approximately equal to 5.8 L . The difference in RMSE reductions may be attributed to variations in model design or dataset characteristics. Future research should enhance the predictive capabilities of these models by incorporating additional cow characteristics, such as reproductive data and pasture allocation information [71]. Expanding the dataset to include these variables will provide a more comprehensive understanding of the factors influencing MY, thereby enhancing the precision of individual cow MY predictions. This advancement is crucial for optimising dairy management practices and maximising overall herd productivity.

4.2. Feed optimisation

Previous research has affirmed that the use of the embedded machine learning model significantly reduces prediction uncertainty compared to traditional methods [72,73]. Without these embedded trained machine learning prediction models, predictions are likely to suffer from a high level of uncertainty, potentially leading to suboptimal feeding strategies. Additionally, we found that using the Scipy library for optimisation are

more efficient and accurate for this type of optimisation problem, as they can handle complex, non-linear relationships inherent in the data and are time efficient as affirmed by [74] compared to linear optimisation approaches.

The substantial increase in total daily concentrate, which rose by approximately 80 kg, leading to a significant rise in the predicted total MY, peaking at 2528.5 kg/day around mid-end of May as opposed to the average predicted a total MY of 2367.03 ± 84.48 kg/day over the 91-day period as observed in Fig. 5 highlight the sensitivity of the model to variations in concentrate levels and its capability to adjust to different feeding scenarios, reinforcing the critical role of concentrate allocation in maximising MY. For instance, during periods of higher concentrate allocations, such as between mid-end of May, the model demonstrated a higher prediction accuracy, improving by up to ~ 23 %. Conversely, when the differential evolution optimisation encountered constraints, it adapted by refining the most feasible solution, achieving the highest optimised MY of 2528.5 kg.day^{-1} within a set daily concentrate limit (Fig. 5). These results suggest that even minor adjustments in concentrate levels can significantly impact milk production, particularly during periods of low yield. The ability of the model to accurately predict these variations shows its potential for optimising dairy production. Further research may explore the underlying mechanisms driving these responses and the long-term sustainability of this optimisation strategy.

4.3. *Individual feeding strategies: data-driven optimisation vs. flat rate allocation*

Our study examined the comparative effectiveness of data-driven optimisation strategies for tailoring grain-based concentrate levels in dairy cows, contrasting them with traditional flat rate allocation methods. The findings highlight substantial potential advantages in MY optimisation and feed efficiency when feeding strategies are tailored to the specific characteristics and behaviour of each cow. Implementing data-driven optimisation models that integrate machine learning techniques has shown significant improvements in MY outcomes.

In contrast, conventional flat rate allocation methods, while simplistic, disregard the individual variabilities among cows. This oversight can result in inefficient feed utilisation, posing risks of both overfeeding and underfeeding. Additionally, farmers often struggle to determine optimal daily supplement allocation without manual intervention. For instance, they might intuitively allocate 6–7 kg DM of supplement to high-milk-yielding cows and cows in early lactation, decreasing by 0.5 kg DM for every 5-litre drop in milk production. It was reported that this decision-making rule is favoured by farmers due to its simplicity and practicality. Over time, farmers reported no significant economic gain or increase in MY from individual feeding strategies, despite many using group-based feeding rather than the individual approach studied by others [8,75,76].

Hence, this limitation emphasises the necessity to transition from static feeding regimes in dairy farming. Our study emphasises that the 8 % theoretical potential MY improvement observed aligns favourably with the findings reported by García et al [9], whose experimental results highlighted that feeding concentrates based on individual cow requirements consistently increased milk solids yield compared to fixed-rate allocations across the herd. In their grazing experiment with 50 lactating Holstein–Friesian dairy cows, cows fed according to individual needs reported an increase in MY and milk fat yield by 3.0 % and 11.1 %, respectively, leading to a significant 7.0 % rise in total milk solids yield. This substantiates the importance of tailoring feed strategies to individual cow characteristics and behaviours to optimise dairy farm productivity effectively [77].

The absence of strong correlations between CONC allocation adjustments and traditional cow characteristics challenges conventional approaches to individualised feeding. While dairy farmers typically base feeding decisions on easily observable traits such as MY, LW, or SOL, our findings suggest that optimal allocation patterns emerge from complex interactions that transcend these simple relationships. This supports ML approaches, where algorithms can identify patterns in high-dimensional data that are invisible to traditional statistical methods. The systematic over-allocation identified by the optimisation process raises important questions about current feeding practices in pasture-based systems. Rather than simply reflecting algorithmic efficiency, these results may indicate fundamental limitations in how CONC allocation decisions are currently made on commercial farms. The larger reductions observed in early and late lactation cows compared to mid-lactation animals may reflect differences in metabolic efficiency across lactation stages, as feed conversion efficiency has been shown to vary throughout the lactation cycle with peak efficiency typically occurring during mid-lactation when cows are at their highest production potential [78]. These findings have broader implications for the adoption of precision feeding technologies in the dairy industry. The weak individual-level correlations, combined with significant group-level differences, suggest that effective precision feeding systems must operate at multiple scales simultaneously - accounting for both individual variation and broader biological patterns. This complexity may explain why simple rule-based feeding systems, despite their apparent logic, fail to capture the full potential for feed efficiency improvements in commercial dairy operations.

Another major downside of the traditional method for feed optimisation is that manually adjusting daily individual cow feed is time-consuming. For large herds on pasture-based dairy systems, this time constraint makes it difficult for farmers to consistently follow feeding rules [79,80]. However, with the availability of data and the implementation of optimisation algorithms, as exhibited in our study, these processes can become significantly less time-consuming and more feasible for farms with large herds. Nevertheless, our study encountered several notable limitations. Firstly, the predictive accuracy of the model was hindered by insufficient training data per cow, limiting its ability to effectively learn from individual historical data during evaluation. Addressing this limitation in future research would involve increasing the volume of individual cow-level training data to enhance predictive capabilities. Secondly, the relatively narrow and fixed concentrate allocations (8–10 kg/cow/day) applied during the study period significantly limited the ability of the model to explore and learn from a broader range of feeding strategies. This constraint reduced the diversity of feeding scenarios available for analysis, thereby limiting the capacity of the model to generalise and provide reliable recommendations beyond the observed data range. These constraints occasionally struggled to manage sudden spikes in total actual concentrate levels on certain days. For instance, on days when the total amount of actual concentrate fed to cows exceeded 732 kg, the model struggled to converge.

Fig. 8 illustrates instances where the predictions of the model deviated excessively (up to 31.3 % increase), notably when upper bounds of 9 kg/cow.day^{-1} were extended to 9.5 kg/cow.day^{-1}, resulting in inaccurate predictions. This constraint particularly impacted the evaluation of sensitivity of the cow to higher concentrate levels exceeding 9 kg/cow.day^{-1} as opposed to the consistent flow observed in the optimisation process in Fig. 5 when the constraint of concentrate allocations was at the maximum value of 9kg/cow.day^{-1}, reducing the risk of abrupt fluctuations in MY. This suggests that maintaining a controlled range for concentrate allocation results helps to provide more stable improvements in MY predictions.

From our study, adopting data-driven optimisation models proves potentially advantageous for modern dairy farms equipped with automated individual feeders. These systems enable real-time monitoring and adjustment of feeding patterns based on the specific nutritional requirements of each cow and behavioural responses. Our optimisation model offers a proactive approach to dairy management by continuously learning from historical feeding data, promoting efficiency, productivity, and animal welfare.

Future research should focus on expanding datasets with varied

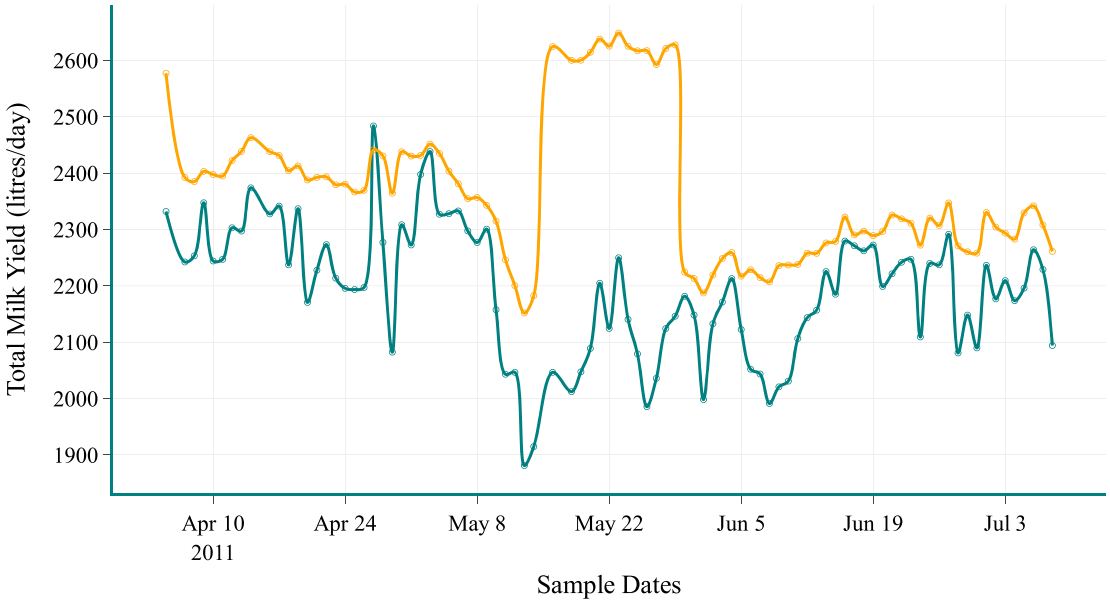


Fig. 8. Total daily milk yield of 81 cows before (green line) and after (orange line) optimisation for the 81 cows across 91 days. All constraints applied at 98-100 % range for total actual concentrate consumed and concentrate levels between 5 and 9.5kg/cow/day.

concentrate amounts to enhance the capacity of the model for dynamic predictions. While refining constraint parameters in optimisation models is vital for minimising computational costs and improving solution optimality, it is important to acknowledge that these adjustments may not yield immediate results in all scenarios. Addressing these challenges will significantly advance the applicability and robustness of data-driven optimisation approaches in dairy farming. Integrating refined optimisation strategies, such as those provided by Scipy [81], Pyomo, Gurobi, and CPLEX, can optimise MY and improve feed efficiency, but their effectiveness will depend on careful implementation and validation within specific farming contexts. By leveraging advanced machine learning techniques, dairy farmers can tailor feeding strategies to individual cow characteristics, thereby maximising profitability while promoting cattle health and welfare. This approach represents a significant step forward in modernising dairy operations to meet industry demands while ensuring long-term sustainability, although further exploration is needed to fully understand the practical implications and challenges of these technologies.

4.4. Limitations and implications for interpretation

A fundamental limitation of this study is the nature of the dataset used for model development and optimisation. The concentrate allocation data (8–10 kg/cow/day) was not experimentally manipulated but rather reflected management decisions based on existing milk production levels (<30 L/d: 8 kg; 30 to 35 L/d: 9 kg; >35 L/d: 10 kg). This approach describes correlations under standard practice rather than direct cause-and-effect responses to varied concentrate levels. Consequently, our ML models learned from associations between concentrate allocation and milk yield within existing management frameworks, rather than from controlled experimental data showing how individual cows respond to changes in concentrate allowance. The optimisation results therefore represent theoretical potential based on these observed correlations, not validated cause-and-effect relationships. This limitation is critical for interpreting our 8 % milk yield improvement, which should be viewed as a theoretical simulation rather than a proven response to individualised feeding strategies. Future research should validate these findings through controlled experiments where concentrate levels are systematically varied for individual cows to establish true dose-response relationships and confirm the practical efficacy of data-driven individualised feeding approaches.

5. Conclusions

We developed an optimisation model to tailor grain-based concentrate levels to individual cows using comprehensive cow-related information. The study aimed to devise a data-driven optimisation model, integrated with a machine learning model that learns individual cows feeding routine and individually distributes grain-based concentrate to individual dairy cows in a more profitable way to maximise the total milk yield for the herd. By embedding a machine learning algorithm (the pre-trained Random Forest model) into the optimisation process, our method leverages the learned functions from the trained model, using animal characteristics to accurately predict individual milk yield. This integration helps create a connection between the intrinsic behaviour of individual cows and their feeding patterns.

Among the sixteen evaluated models utilising a 5-fold cross-validation technique, the Random Forest model emerged as the most suitable predictive model for milk yield modelling in the study. Additionally, our study provided valuable insights into the importance of individual cow variability, demonstrating that accounting for these differences in cow-specific traits is crucial for improving milk yield modelling and for optimisation. Linking the Random Forest predictions at individual cow level with a non-linear optimisation method, the Differential Evolution algorithm resulted in a theoretical mean herd level increase in milk yield of 8.11 % over 91 days, without increasing the total amount of concentrate ‘fed’ to the cows.

Despite these promising results, certain limitations were encountered due to constraints within the existing dataset. The dataset reflects management decisions based on existing production levels rather than experimental manipulation of concentrate allowance, describing correlations under standard practice rather than direct cause-and-effect responses. Insufficient training data per cow limited the ability of the model to learn effectively from individual historical data, while fixed feeding rates restricted the exploration of diverse feeding options. Additionally, the optimisation constraints applied to the model require thorough evaluation to enhance computational efficiency, as they often struggle to manage sudden spikes in total actual concentrate levels, necessitating improved strategies to mitigate costs and operational bottlenecks. These limitations present opportunities for refinement in

future research endeavours.

Ultimately, data-driven approaches for individual cow feed optimisation hold promise for maximising daily total milk yield while maintaining consistent feed levels. Further validation through field trials with systematic concentrate manipulation or actual dairy farm applications will be crucial to confirm their effectiveness in real-world scenarios. Such methodologies offer potential benefits for enhancing feed and pasture management within dairy systems.

Funding statement

This research was funded by the Dairy UP program, a collaborative RD&E program for New South Wales, Australia (www.dairyup.com.au) through the academic scholarship awarded to Blessing N. Azubuike.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

If this manuscript involves research on animals or humans, it is imperative to disclose all approval details.

If Yes, please provide your text here:

This study was conducted in accordance with the ethical standards outlined by the relevant institutional and national guidelines. The research involving animals was reviewed and approved by The University of Sydney's Animal Ethics Committee (Protocol Number: 2019/1608). All procedures performed in this study adhered to ethical principles, ensuring the welfare of the animals involved. The study complied with the Australian Code for the Care and Use of Animals for Scientific Purposes, 8th Edition (2013).

CRediT authorship contribution statement

Blessing Nnenna Azubuike: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Chlingaryan:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Martin Correa-Luna:** Writing – review & editing, Supervision, Conceptualization. **Cameron E.F. Clark:** Writing – review & editing, Investigation, Conceptualization. **Sergio C. Garcia:** Writing – review & editing, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was supported by the research scholarship granted by the DairyUP project at the University of Sydney.

Appendix

Table A
1. Summary statistics of cows characteristics, concentrate consumed and milk yield by the number of lactations and stages of lactation.

Number of Lactations																
Metrics	1				2				3				4			
	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)
Mean	161.90	514.70	8.09	24.11	161.70	586.05	8.30	29.13	165.31	626.49	8.36	30.61	161.72	660.18	8.42	30.60
Std	79.72	45.19	0.30	4.76	80.63	49.59	0.63	6.60	79.93	45.59	0.62	5.80	80.28	50.90	0.70	6.65
Minimum	7.00	401.00	8.00	8.00	9.00	455.00	8.00	8.00	9.00	504.00	8.00	8.80	8.00	312.00	8.00	8.10
25 %	94.00	481.00	8.00	20.90	94.00	551.00	8.00	25.20	97.00	600.50	8.00	26.80	93.25	633.00	8.00	26.30
Median	161.00	512.00	8.00	23.90	162.00	586.00	8.00	28.80	166.00	625.00	8.00	30.50	161.00	666.00	8.00	30.60
75 %	230.00	546.00	8.00	27.20	230.00	620.00	8.00	32.90	234.00	657.00	9.00	34.20	230.00	691.00	9.00	34.90
Maximum	306.00	798.00	10.00	42.30	310.00	886.00	10.00	55.30	308.00	771.00	10.00	55.40	308.00	823.00	10.00	53.40

Stages of Lactation												
Metrics	Early Lactation				Mid-Lactation				Late Lactation			
	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)	DIM (days)	LW (kg/cow)	CONC (kg/cow)	MY (L/cow)
Mean	61.63	567.13	8.53	31.48	150.49	576.76	8.23	27.81	250.99	611.69	8.10	25.65
Std	23.96	78.45	0.75	6.79	28.78	72.19	0.52	5.97	29.47	68.41	0.34	5.75
Minimum	7.00	312.00	8.00	8.90	101.00	401.00	8.00	8.00	201.00	439.00	8.00	8.00
25 %	44.00	501.00	8.00	26.70	126.00	514.00	8.00	23.80	225.00	555.00	8.00	23.80
Median	63.00	569.50	8.00	31.20	151.00	578.00	8.00	27.70	251.00	610.00	8.00	25.60
75 %	82.00	628.00	9.00	35.90	175.00	635.00	8.00	31.70	276.00	664.75	8.00	29.40
Maximum	100.00	886.00	10.00	55.40	200.00	798.00	10.00	52.00	310.00	823.00	10.00	52.90

Notes: DIM = Days in Milk (days), LW = Liveweight (kg/cow/day), CONC = Concentrate level consumed (kg/cow/day), MY = Milk Yield (L/cow/day.

Data availability

Data will be made available on request.

References

[1] I. Rugoho, C.J.P. Gourley, M.C. Hannah, Nutritive characteristics, mineral concentrations and dietary cation–anion difference of feeds used within grazing-based dairy farms in Australia, Anim. Prod. Sci. 57 (5) (2017), https://doi.org/10.1071/an15761.

[2] C.E.F. Clark, R. Kaur, L.O. Millapan, H.M. Golder, P.C. Thomson, A. Horadagoda, M.R. Islam, K.L. Kerrisk, S.C. Garcia, The effect of temperate or tropical pasture grazing state and grain-based concentrate allocation on dairy cattle production and behavior, J. Dairy Sci. 101 (6) (2018) 5454–5465, https://doi.org/10.3168/jds.2017-13388.
[3] E.S. Kolver, Nutritional limitations to increased production on pasture-based systems, Proceed. Nutrit. Soc. 62 (2) (2003) 291–300, https://doi.org/10.1079/PNS2002200.
[4] M.F. Dida, S.C. Garcia, L.A. Gonzalez, Dietary concentrate supplementation increases milk production and reduces predicted greenhouse gas emission intensity in pasture-based commercial dairy farms, J. Dairy Sci. 107 (8) (2024) 5639–5652, https://doi.org/10.3168/jds.2023-24303.
[5] E. Finneran, P. Crosson, P. O'Kiely, L. Shalloo, D. Forristal, M. Wallace, Stochastic simulation of the cost of home-produced feeds for ruminant livestock systems, J. Agri. Sci. 150 (1) (2012) 123–139, https://doi.org/10.1017/S002185961100061X.
[6] P.J. Purcell, A.J. Dale, A.W. Gordon, J. Barley, C.P. Ferris, Effects of predicted milk yields sustained by grazed grass on dairy cow performance and concentrate requirements throughout the grazing season, Grass Forage Sci. 71 (3) (2016) 389–402, https://doi.org/10.1111/gfs.12193.
[7] A.-L. Craig, A.W. Gordon, G. Hamill, C.P. Ferris, Milk composition and production efficiency within feed-to-yield systems on commercial dairy farms in Northern Ireland, Animals 12 (14) (2022) 1771, https://doi.org/10.3390/ani12141771.
[8] B.T. Dela Rue, C.R Eastwood, Individualised feeding of concentrate supplement in pasture-based dairy systems: practices and perceptions of New Zealand dairy farmers and their advisors, Anim. Prod. Sci. 57 (7) (2017) 1543–1549, https://doi.org/10.1071/AN16471.
[9] S.C. García, M. Pedernera, W.J. Fulkerson, A. Horadagoda, K. Nandra, Feeding concentrates based on individual cow requirements improves the yield of milk solids in dairy cows grazing restricted pasture, Aust. J. Exp. Agri. 47 (5) (2007) 502–508, https://doi.org/10.1071/EA05349.
[10] E.L. Brady, E.T. Kelly, M.B. Lynch, A.G. Fahey, K.M. Pierce, F.J. Mulligan, The effect of concentrate feeding strategy and dairy cow genotype on milk production, pasture intake, body condition score and metabolic status under restricted grazing conditions. Livestock Science, 2022, p. 256, https://doi.org/10.1016/j.livsci.2021.104815.
[11] J.C.S. Henriksen, M.R. Weisbjerg, P. Lovendahl, T. Kristensen, L. Munksgaard, Effects of an individual cow concentrate strategy on production and behavior, J. Dairy Sci. 102 (3) (2019) 2155–2172, https://doi.org/10.3168/jds.2018-15477.
[12] A.J. Dale, S. McGettrick, A.W. Gordon, C.P. Ferris, The effect of two contrasting concentrate allocation strategies on the performance of grazing dairy cows, Grass Forage Sci. 71 (3) (2016) 379–388, https://doi.org/10.1111/gfs.12185.
[13] G. André, P.B.M. Berentsen, G. Van Duinkerken, B. Engel, A.G.J.M.O. Lansink, Economic potential of individual variation in milk yield response to concentrate intake of dairy cows, J. Agric. Sci. 148 (3) (2010) 263–276, https://doi.org/10.1017/S0021859610000134.
[14] D.C. Lawrence, M. O'Donovan, T.M. Boland, E. Lewis, E. Kennedy, The effect of concentrate feeding amount and feeding strategy on milk production, dry matter intake, and energy partitioning of autumn-calving Holstein-Friesian cows, J. Dairy Sci. 98 (1) (2015) 338–348, https://doi.org/10.3168/jds.2014-7905.
[15] D.M. Liebe, R.R. White, Analytics in sustainable precision animal nutrition, Animal Frontiers: Rev. Magazine Animal Agri. 9 (2) (2019) 16–24, https://doi.org/10.1093/af/vfz003.
[16] M.W. Little, N.E. O'Connell, C.P. Ferris, A comparison of individual cow versus group concentrate allocation strategies on dry matter intake, milk production, tissue changes, and fertility of Holstein-Friesian cows offered a grass silage diet, J. Dairy Sci. 99 (6) (2016) 4360–4373, https://doi.org/10.3168/jds.2015-10441.
[17] V.K. Chauhan, Stochastic Optimization for Large-scale Machine Learning, 1st Edition ed, CRC Press, 2021, https://doi.org/10.1201/9781003240167.
[18] S. Balhara, A.K. Balhara, N. Dahiya, Himanshu, R. Singh, A. Ruhil, Machine learning algorithms for predicting peak yield in buffaloes using linear traits, Indian J. Anim. Sci. 92 (8) (2022) 1013–1019, https://doi.org/10.56093/ijans.v92i8.122008.
[19] G.M. Dallago, D.M.d. Figueiredo, P.C.d.R. Andrade, R.A.d. Santos, R. Lacroix, D. E. Santschi, D.M. Lefebvre, Predicting first test day milk yield of dairy heifers, Comp. Electro. Agri. 166 (2019) 105032, https://doi.org/10.1016/j.compag.2019.105032.
[20] M. Frizzarin, I.C. Gormley, D.P. Berry, T.B. Murphy, A. Casa, A. Lynch, S. McParland, Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods, J. Dairy Sci. 104 (7) (2021) 7438–7447, https://doi.org/10.3168/jds.2020-19576.
[21] B. Ji, T. Banhazi, C.J.C. Phillips, C. Wang, B. Li, A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm, Biosyst. Eng. 216 (9) (2022) 186–197, https://doi.org/10.1016/j.biosystemseng.2022.02.013.
[22] Q.T. Nguyen, R. Fouchereau, E. Frénod, C. Gerard, V. Sincholle, Comparison of forecast models of production of dairy cows combining animal and diet parameters, Comp. Electro. Agri. 170 (2020) 105258, https://doi.org/10.1016/j.compag.2020.105258.
[23] J.L. Ellis, M. Jacobs, J. Dijkstra, H. van Laar, J.P. Cant, D. Tulpan, N. Ferguson, Review: synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data, Animal 14 (2) (2020) s223–s237, https://doi.org/10.1017/S1751731120000312.
[24] July 4-6Chair C.E.F Clark, What are we feeding our cows? in: K. Kerrisk, S. C. García, S. Catt, M. Heward (Eds.), *Current Topics in Dairy Production.* [Symposium], *Current Topics in Dairy Production.*[Symposium], 18 Dairy Research Foundation, University of Sydney: Sydney, 2013, pp. 44–49, https://dairy-foundation.sydney.edu.au/wp-content/uploads/2019/10/2013-PROCEEDINGS-Kiama.pdf.
[25] L. Hanrahan, N. McHugh, T. Hennessy, B. Moran, R. Kearney, M. Wallace, L. Shalloo, Factors associated with profitability in pasture-based systems of milk production, J. Dairy Sci. 101 (6) (2018) 5474–5485, https://doi.org/10.3168/jds.2017-13223.
[26] E. Ruelle, L. Delaby, M. Wallace, L. Shalloo, Using models to establish the financially optimum strategy for Irish dairy farms, J. Dairy Sci. 101 (1) (2018) 614–623, https://doi.org/10.3168/jds.2017-12948.
[27] F. Galton, Regression towards mediocrity in hereditary stature, J. Anthropo. Inst. Great Britain Ireland 15 (1886) 246–263, https://doi.org/10.2307/2841583.
[28] K. Kumari, S. Yadav, Linear regression analysis study, J. Practic. Cardiovas. Sci. 4 (1) (2018) 33–36, https://doi.org/10.4103/jpcs.jpcs_8_18.
[29] A. Schneider, G. Hommel, M. Blettner, Linear regression analysis: part 14 of a series on evaluation of scientific publications, Dtsch Arztebl Int. 107 (44) (2010) 776–782, https://doi.org/10.3238/2Farztebl.2010.0776.
[30] J.M. Stanton, Galton, Pearson, and the Peas: a brief history of linear regression for statistics instructors, J. Stat. Edu. 9 (3) (2001) 1–13, https://doi.org/10.1080/10691898.2001.11910537.
[31] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Royal Stat. Soc. Seri. B 58 (1) (1996) 267–288. http://www.jstor.org/stable/2346178.
[32] R. Hoerl, Ridge regression: a historical context, Technometrics 62 (4) (2020) 420–425, https://doi.org/10.1080/00401706.2020.1742207.
[33] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. Royal Stat. Soc. Seri. B 67 (2) (2005) 301–320, https://doi.org/10.1080/00401706.2020.1742207.
[34] E. Ostertagová, Modelling using polynomial regression, Proced. Eng. 48 (2012) 500–506, https://doi.org/10.1016/j.proeng.2012.09.545.
[35] S.M. Stigler, Gergonne's 1815 paper on the design and analysis of polynomial regression experiments, Historia Math. 1 (4) (1974) 431–439, https://doi.org/10.1016/0315-0860(74)90033-0.
[36] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS Regression), WIREs Comput. Stat. 2 (1) (2010) 97–106, https://doi.org/10.1002/wics.51.
[37] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chem. Intell. Lab. Syst. 58 (2) (2001) 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.
[38] A. Ben-Hur, D. Horn, H.T. Siegelmann, V. Vapnik, Support vector clustering, J. Mach. Learn. Res. 2 (2002) 125–137. https://dl.acm.org/doi/10.5555/944790.944807.
[39] C. Cortes, V. Vapnik, Support-Vector Networks, Mach Learn 20 (3) (1995) 273–297, https://doi.org/10.1023/A:1022627411411.
[40] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Annals Stat. 29 (5) (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[41] L. Mason, J. Baxter, P. Bartlett, M. Frean, Boosting algorithms as gradient descent, in: Proceedings of the 12th International Conference on Neural Information Processing Systems, Denver, CO, 1999. https://dl.acm.org/doi/10.5555/3009657.3009730.
[42] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, KDD '16: The 22nd ACM SIGKDD Int. Conference Know. Dis. Data Min. (2016) 785–794, https://doi.org/10.1145/2939672.2939785.
[43] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Proces. Syst. 30 (NIPS) (2017).
[44] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, Adv. Neural Inf. Proces. Syst. 31 (2018), https://doi.org/10.48550/arXiv.1706.09516.
[45] B. Kamiński, M. Jakubczyk, P. Szufel, A framework for sensitivity analysis of decision trees, Centr. European J. Operat. Res. 26 (1) (2018) 135–159, https://doi.org/10.1007/s10100-017-0479-6.
[46] D.v. Winterfeldt, W. Edwards, Decision Analysis and Behavioral Research, Cambridge University Press, 1986, pp. 63–89. ISBN 0-521-27304-8.
[47] L. Breiman, Random forests, Mach. Learn 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[48] A. Liaw, M. Wiener, Classification and regression by randomForest, R news 2 (3) (2002) 18–22. https://journal.r-project.org/articles/RN-2002-022/.
[49] H. Tin Kam, Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 1995, https://doi.org/10.1109/ICDAR.1995.598994.
[50] E. Fix, J.L. Hodges, Discriminatory analysis. Nonparametric discrimination: consistency properties, Int. Stat. Rev. 57 (3) (1989) 238–247, https://doi.org/10.2307/1403797.
[51] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, J.B. Mitchell, Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization, J .Chem. Inf. Model 46 (6) (2006) 2412–2422, https://doi.org/10.1021/ci060149f.
[52] N. Lawrence, M. Seeger, R. Herbrich, *Fast sparse Gaussian process methods: the informative vector machine* [Conference paper]. Advances in Neural Information Processing Systems, 2003. https://www.scopus.com/inward/record.uri?eid=2-s2.0-80053225881&partnerID=40&md5=943790363627bf62e98a5709625f4aac.
[53] C.E. Rasmussen, C.K. Williams, Gaussian Processes For Machine Learning, 1, The MIT Press, Cambridge, Massachusetts, 2006, https://doi.org/10.7551/mitpress/3206.001.0001.

[54] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, S. Aigrain, Gaussian processes for time-series modelling, Philos. Trans. A Math. Phys. Eng. Sci. 371 (1984) (2013) 20110550, https://doi.org/10.1098/rsta.2011.0550.
[55] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions [Article], J. Math. Psychol. 85 (2018) 1–16, https://doi.org/10.1016/j.jmp.2018.03.001.
[56] M. Titsias, Variational learning of inducing variables in sparse gaussian processes, J. Mach. Learn. Res. 5 (2009) 567–574, in: https://proceedings.mlr.press/v5/titsias09a/titsias09a.pdf.
[57] T.H. Tieleman, G, Lecture6.5-RMSProp: divide the gradient by a running average of its recent magnitude, Coursera: Neural Net. Mach. Learn. 4 (2012) 26–31.
[58] L. Bottou, On-line learning and stochastic approximations, in: D. Saad (Ed.), On-Line Learning in Neural Networks, Cambridge University Press, 1999, pp. 9–42, https://doi.org/10.1017/CBO9780511569920.003.
[59] L. Bottou, O. Bousquet, The tradeoffs of large-scale learning, in: S. Sra, S. Nowozin, S.J. Wright (Eds.), Optimization For Machine Learning, The MIT Press, Cambridge, 2011, pp. 351–368, https://doi.org/10.7551/mitpress/8996.003.0015.
[60] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: International Conference on Learning Representations, 2015. http://arxiv.org/abs/1412.6980.
[61] R.E. Franzoi, B.C. Menezes, J.D. Kelly, C.L.E. Swartz, Adaptive least-squares surrogate modeling for reaction systems, in: Y. Yamashita, M. Kano (Eds.), Computer Aided Chemical Engineering, Computer Aided Chemical Engineering, 49, Elsevier, 2022, pp. 1705–1710, https://doi.org/10.1016/B978-0-323-85159-6.50284-0.
[62] Wang, K., Wilder, B., Perrault, A., & Tambe, M. (2020). Automatically learning compact quality-aware surrogates for optimization problems. *ArXiv, abs/2006.10815*. https://doi.org/10.48550/arXiv.2006.10815.
[63] B. Williams, S. Cremaschi, Novel tool for selecting surrogate modeling techniques for surface approximation, in: M. Türkay, R. Gani (Eds.), Computer Aided Chemical Engineering, Computer Aided Chemical Engineering, 50, Elsevier, 2021, pp. 451–456, https://doi.org/10.1016/B978-0-323-88506-5.50071-1.
[64] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogram. Remot. Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.
[65] D.B. Jensen, M. van der Voort, H. Hogeveen, Dynamic forecasting of individual cow milk yield in automatic milking systems, J. dairy sci. 101 (11) (2018) 10428–10439, https://doi.org/10.3168/jds.2017-14134.
[66] C. O'Leary, C. Lynch, An Evaluation of Machine Learning Approaches for Milk Volume Prediction in Ireland. *2022*. https://doi.org/10.1109/issc55427.2022.9826160.
[67] M.D. Murphy, M.D. Murphy, M.J. O'Mahony, L. Shalloo, P. French, J. Upton, Comparison of modelling techniques for milk-production forecasting, J. Dairy Sci. 97 (6) (2014) 3352–3363. 10.3168/jds.2013-7451.
[68] H. Shu, Y. Li, J. Bindelle, Z. Jin, T. Fang, M. Xing, L. Guo, W. Wang, Predicting physiological responses of dairy cows using comprehensive variables, Comp. Electron. Agri. 207 (2023) 107752, https://doi.org/10.1016/j.compag.2023.107752.
[69] G.-J. Streefland, F. Herrema, M. Martini, A Gradient boosting model to predict the milk production, Smart Agri. Tech. 6 (2023) 100302, https://doi.org/10.1016/j.atech.2023.100302.
[70] M. Salamone, I. Adriaens, A. Vervaet, G. Opsomer, H. Atashi, V. Fievez, B. Aernouts, M. Hostens, Prediction of first test day milk yield using historical records in dairy cows, Animal 16 (11) (2022) 100658, https://doi.org/10.1016/j.animal.2022.100658.
[71] P.R. Shorten, Computer vision and weigh scale-based prediction of milk yield and udder traits for individual cows, Comp. Electron. Agric. 188 (6) (2021) 106364, https://doi.org/10.1016/j.compag.2021.106364.
[72] F. Agakov, E. Bonilla, J. Cavazos, B. Franke, G. Fursin, M.F.P. O'Boyle, J. Thomson, M. Toussaint, C.K.I. Williams, Using machine learning to focus iterative optimization, in: Proceedings of the International Symposium on Code Generation and Optimization, 2006, https://doi.org/10.1109/CGO.2006.37.
[73] A.M. Schweidtmann, D. Bongartz, A. Mitsos, Optimization with trained machine learning models embedded, in: P.M. Pardalos, O.A. Prokopyev (Eds.), Encyclopedia of Optimization, Springer International Publishing, 2020, pp. 1–8, https://doi.org/10.1007/978-3-030-54621-2_735-1.
[74] L.M. Campos, H. Ringer, M. Chung, M.D. Hanigan, Application of a mathematical framework for the optimization of precision-fed dairy cattle diets, Animal 17 (2023) 101001, https://doi.org/10.1016/j.animal.2023.101001.
[75] S. García, M. Islam, C. Clark, P. Martin, Kikuyu-based pasture for dairy production: a review, Crop. Pasture Sci. 65 (8) (2014) 787–797, https://doi.org/10.1071/CP13414.
[76] J.L. Hills, S.C. García, B. Dela Rue, C.E.F. Clark, Limitations and potential for individualised feeding of concentrate supplements to grazing dairy cows, Anim. Prod Sci. 55 (7) (2015) 922–930, https://doi.org/10.1071/AN14855.
[77] P.J.M. Raedts, J.L. Hills, Milk yield and feeding behaviour responses to two flat-rate levels of concentrate supplementation fed over a period of 8 months to cohorts of grazing dairy cows, differing in genotype, bodyweight, or milk yield, Anim . Prod. Sci. 64 (7) (2024), https://doi.org/10.1071/AN23142.
[78] A.M. Hurley, N. Lopez-Villalobos, S. McParland, E. Lewis, E. Kennedy, M. O'Donovan, J.L. Burke, D.P. Berry, Characteristics of feed efficiency within and across lactation in dairy cows and the effect of genetic selection, J. Dairy Sci. 101 (2) (2018) 1267–1280, https://doi.org/10.3168/jds.2017-12841.
[79] L. Kota, K. Jarmai, Improving optimization using adaptive algorithms, Pollack Periodic.: An Int. J. Eng. Infor. Sci. 16 (1) (2021) 14–18, https://doi.org/10.1556/606.2020.00180.
[80] Solanki, M.S. (2021). A Brief description on Optimization Techniques: a virtual international conference redefining and transforming role higher education in sustainable development (Conference paper). 2, 125-132. https://doi.org/10.46402/2021.02.25.
[81] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. SciPy, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (3) (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.

CHAPTER 4

Leveraging artificial intelligence and evolutionary algorithms for optimising cow supplementation and milk production

Chapter 3 demonstrated precision feeding potential using flat-rate data. This chapter advances this with real world complexity. We leverage a massive commercial dataset comprising over 450 thousand records that explicitly captured individual cow specific concentrate rates, revealing far greater feeding variability than previously examined in Chapter 3. Integrating advanced artificial intelligence and multi-objective evolutionary algorithms, we optimise concentrate allocation while incorporating environmental variables to address the complexity and variability of large scale commercial pasture systems. This framework achieved statistically validated increases in total milk yield ranging from 6.63% to 8.64% across ten independent algorithm runs without increasing the overall feed budget, with NSGA-II demonstrating the highest performance (8.64%) and fastest computation time, validating the scalability and commercial viability of precision allocation demonstrated in Chapter 3. All code implementations described in this chapter are openly available in the GitHub repository at https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-4-machine-learning. Key code excerpts are presented in Appendix B for reference.

Artificial Intelligence in Agriculture (2026) 16(2), pp 926-939

PUBLISHED MANUSCRIPT

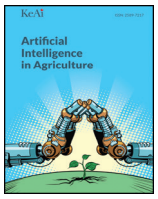
The published version of this manuscript is included on the following page.

Artificial Intelligence in Agriculture 16 (2026) 926–939

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Research Paper

Leveraging artificial intelligence and evolutionary algorithms for optimising cow supplementation and milk production

Blessing Nnenna Azubuike [a,d,*], Anna Chlingaryan [b,d], Martin Correa-Luna [a,d], Cameron E.F. Clark [c,d], Sergio C. Garcia [a,d]

[a] Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW 2570, Australia
[b] Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Camden 2570, Australia
[c] Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW 2650, Australia
[d] Dairy UP Program, Camden, NSW 2570, Australia

ARTICLE INFO

Article history:
Received 1 June 2025
Received in revised form 31 March 2026
Accepted 7 April 2026
Available online 08 April 2026

Keywords:
Dairy cattle
Evolutionary algorithms
Feed optimisation
Individual concentrate allocation
Machine learning

ABSTRACT

Efficient allocation of grain-based concentrate is essential for maximising milk yield and improving profitability in dairy farming. This study optimised concentrate allocation for dairy cows by integrating machine learning with evolutionary algorithms to enhance milk yield. Data from 1105 cows were analysed to model milk yield using predictor variables including breed, days in milk, stage of lactation, lactation number, daily concentrate allocation and climatic factors. For optimisation, records for 165 cows within a 30-day window were identified using a Genetic Algorithm. Three machine learning models (Gradient Boosting Machines, Extreme Gradient Boosting and Random Forest) were evaluated, with GBM showing superior predictive accuracy ($R^2 = 0.75$ training, 0.61 testing; RMSE = 3.02 L/cow·day^{-1} training, 3.47 L/cow·day^{-1} testing). The pre-trained GBM was integrated into four evolutionary algorithms (NSGA-II, SPEA-II, SMS-EMOA and RVEA) to simultaneously maximise herd-level daily milk yield and minimise deviations in daily concentrate allocation. Concentrate levels per cow were constrained between 6 and 11 kg per day, with adaptive bounds allowing no more than a 2 kg change per cow per day from the previous allocation. Without increasing herd-level supplementation allowance, statistical analysis based on 10 independent runs per algorithm revealed NSGA-II achieved the highest mean milk yield increase of 8.64 ± 0.10%, significantly outperforming SPEA-II (7.94 ± 0.13%), RVEA (7.41 ± 0.10%), and SMS-EMOA (6.63 ± 0.24%). One-way ANOVA confirmed significant differences among all algorithms ($F = 303.24$, $p < 0.001$), with NSGA-II also demonstrating superior computational efficiency (74.9 ± 0.7 s per day). These findings demonstrate the potential of combining machine learning and evolutionary algorithms to optimise feed use efficiency, highlighting promising practical applications of individualised feeding approaches to enhance milk production in dairy systems.

1. Introduction

Efficient feed allocation is pivotal in dairy farming, directly influencing profitability and milk production. In pasture-based dairy systems, grain-based concentrates, rich in energy, are commonly used to enhance milk yield (MY) and supplement pasture intake; however, they significantly contribute to operational costs, often comprising up to 60% of total production expenses (Craig et al., 2022; Uyeh et al., 2018). Traditional feeding practices in these systems typically employ a flat-rate approach, distributing concentrates uniformly across the herd without accounting for individual cow variations such as breed, lactation stage, days in milk (DIM), and historical production performance. This homogenised method can lead to feed inefficiencies and inconsistent MY, ultimately affecting the economic viability of the farm (Dela Rue and Eastwood, 2017; Hills et al., 2015).

Optimising feed allocation remains a complex challenge despite advancements in agricultural technology. Current strategies often fail to utilise the extensive data generated on modern farms (Cockburn, 2020) or adapt to the dynamic nutritional requirements of individual cows (Dela Rue and Eastwood, 2017; Hills et al., 2015). This results in suboptimal feeding regimes that do not fully exploit the potential of each cow for milk production, leading to underutilisation of resources and limited productivity gains. Innovative approaches are needed to process complex datasets, informing feeding decisions and tailoring concentrate allocation to maximise MY while maintaining cost-effectiveness (Brady et al., 2022; Dale et al., 2016; Henriksen et al., 2019).

* Corresponding author at: 380 Werombi Road, Centre for Carbon, Water and Food (CCWF), Faculty of Science, University of Sydney, Camden, NSW 2570, Australia.
E-mail address: blessing.azubuike@sydney.edu.au (B.N. Azubuike).

https://doi.org/10.1016/j.aiia.2026.04.003

Machine learning (ML) techniques and Artificial Intelligence (AI) offer a promising solution by enabling the analysis of large datasets to uncover intricate patterns between cow-specific factors and MY (Cockburn, 2020). By developing predictive models based on variables such as breed; lactation number; DIM; and daily concentrate intake; farmers can customise feed allocation. Additionally; evolutionary algorithms provide robust methods for multi-objective optimisation; balancing goals like maximising MY and minimising feed costs (Ji et al., 2022; Nguyen et al., 2020; Shu et al., 2023; Souza et al., 2022; Streefland et al., 2023).

Previous research has explored diet optimisation and cost minimisation in dairy farming using mathematical programming techniques, including linear and non-linear programming and goal programming for animal diet formulation (Campos et al., 2023; Das et al., 2023; Notte et al., 2020; Usigbe et al., 2023; Uyeh et al., 2019). Evolutionary algorithms such as genetic algorithms, simulated annealing, and differential evolution have been applied to feed optimisation problems (Astuti and Raj, 2016). While effective in reducing feed costs and optimising resource allocation; these methods often lack integration with machine learning models that could enhance predictive accuracy and computational efficiency (Notte et al., 2021) and focus on group-level strategies rather than individual cow optimisation; limiting their applicability in personalised feeding regimes. The algorithms (linear programming) used are sometimes constrained in nonlinear search spaces or local optima and are tested on limited real-world applications; affecting the generalisability of the results. Models capable of near real-time prediction and optimisation that can be implemented in pasture-based dairy systems are needed as some studies (Souza et al., 2022) have focused on intensive dairy systems. By integrating ML with evolutionary algorithms, this study builds on previous research work that were impeded by prolonged run times and convergence on local optima in conventional optimisation solvers, offering a comprehensive framework for individual cow level feed allocation that could lead to significant gains in milk production efficiency and improved pasture utilisation. We hypothesise that by optimising feed strategies with ML and evolutionary algorithms, the combined approach will deliver a higher marginal MY response per kilogram of concentrate than traditional biologically driven feeding rules, thereby using the available concentrate more efficiently and, in turn, promoting sustainable pasture management.

This study aims to integrate ML models with evolutionary optimisation algorithms to enhance feed allocation strategies in pasture-based dairy farming by developing a comprehensive decision-support system for dairy farmers. It addresses two key questions: can ML reliably predict MY from individual cow characteristics and feeding patterns? and can embedded evolutionary algorithms effectively optimise grain-based concentrate allocation to maximise daily MY? Furthermore, this research examines whether the optimisation process remains profitable and delivers comparable results in pasture-based commercial farms with individual cow feeding systems, such as computerised bail feeding, where significant variability in feed allocation is observed. Building on previous work where this approach has been applied in flat-rate feeding systems (Azubuike et al., 2024) and a small amount dataset, this study also investigates whether increased variability in feeding data might improve the training and performance of the model, particularly when scaling from small to larger datasets to develop a data-driven framework, using all data available, that integrates these approaches to optimise grain-based concentrate allocation on an individual cow basis, thereby enhancing MY and economic sustainability in large-scale pasture-based dairy farm operations.

2. Materials and methods

2.1. Data

This study utilised a dataset collected from an Australian pasture-based commercial dairy farm over a period of 886 days, ranging from 12 January 2022 to 16 June 2024. The dataset comprised 485,449 records from 1105 individual cows, providing detailed daily information of milk production, feeding patterns, and environmental conditions.

Milk production data included individual cow unique identifier (Cow_ID), breed information (Holstein, Holstein Friesian, Friesian × Jersey, Swedish Red-and-White), calving date, days in milk (DIM, days), stage of lactation (SOL, categorical: early, mid, late), number of lactation (NL, categorical: 1, 2, 3, 4), and daily milk yield (MY, L cow^{-1} day^{-1}). Feeding data encompassed daily concentrate intake (CONC, kg cow^{-1} day^{-1}) and the concentrate allocation recommended by Delpro software (DeLaval, Sweden) (drat, kg cow^{-1} day^{-1}), which operates based on predefined feeding rules set by the farm managers. Concentrate was allocated on a dry matter (DM) basis. The grain-based concentrate pellets fed to the cows had an approximate composition of 17% crude protein (CP), 20% neutral detergent fibre (NDF), and 12.5 MJ ME/kg DM, typical of commercial concentrate formulations used in Australian pasture-based dairy systems. Environmental conditions were recorded through minimum and maximum temperatures (min_temp and max_temp, °C), vapour pressure (vp, kPa) and relative humidity levels at minimum and maximum temperatures (rh_tmin and rh_tmax, %). Additional computed variables included expected MY projected by the Delpro software (dexp_my) and average MY over the previous seven days (AvgYieldPrev7d). The stages of lactation were categorised as early (1 to 100 DIM), mid (101 to 199 DIM), and late lactation (200 or more DIM) stages. This classification was essential for analysing milk production patterns relative to different lactation stages.

To ensure the integrity and quality of the data for subsequent analysis, a thorough data preprocessing procedure was implemented. Categorical variables (Cow_ID, breed, SOL, NL) were converted to factor variables and underrepresented groups (breeds with 6000 records or less) were excluded from the dataset to ensure that the models were trained on breeds with adequate representation. Outlier detection and removal were conducted on selected continuous variables: DIM, MY, and daily concentrate intake (CONC). Outliers were identified using the interquartile range (IQR) method, where data points lower than 1.5 times the IQR below the first quartile or higher than 1.5 times the IQR above the third quartile (only those biologically impossible or unlikely values) were discarded. The preprocessing steps reduced the dataset by 6.05%, resulting in a refined dataset (456,078 records for 1053 cows) suitable for accurate and reliable modelling. All data preprocessing and descriptive analyses were conducted using the open-source statistical software R, version 4.4.1.

2.2. Exploratory data analysis

An exploratory data analysis was undertaken to comprehend the underlying patterns and relationships within the dataset before modelling. Spearman rank correlation was employed to examine associations between and among the variables (MY, DIM, concentrate intake, and environmental factors). This non-parametric method was selected to detect monotonic relationships between the variables.

To identify potential groupings among cows, K-means clustering (Bock, 2007) was applied with $k = 2$ clusters, determined as mathematically optimal through comprehensive validation. The optimal number of clusters was identified using three complementary validation metrics: the elbow method, silhouette score (0.494–0.523), and Davies-Bouldin Index (<0.76) (see Supplementary Fig. S1). The clustering algorithm focused on the average MY and average concentrate intake per cow as the variables for grouping. The clustering process was performed separately for each number of lactations to account for variations in milk production patterns across different lactations within the herd with respect to their respective concentrate intake levels, with $k = 2$ consistently identified as optimal across all lactation numbers. Visual representations, such as scatter plots and box plots, were employed to illustrate the relationships between key variables (average MY versus CONC) within each cluster and the outcomes of the clustering analysis.

2.3. Milk yield modelling

The predictive models for MY were developed, evaluated, and optimised using Python (version 3.11.4). Machine learning algorithms (GBM, XGBoost, RF) were implemented using scikit-learn version 1.7.2 (Pedregosa et al., 2012); while the evolutionary algorithms (NSGA-II; SPEA-II; SMS-EMOA; RVEA) for multi-objective optimisation were implemented using Pymoo version 0.6.1.1 (Blank and Deb, 2020). Data manipulation and preprocessing were performed using pandas version 2.2.3 and NumPy (version 1.26.4). Following the exploratory data analysis, three machine learning algorithms were employed to predict daily MY for individual cows: Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), and Random Forest (RF). The target variable for prediction was MY and the predictor variables included both categorical (breed, SOL, NL) and continuous variables (DIM, CONC, min_temp, max_temp, rh_tmin, rh_tmax, and `vp`). DIM and SOL were both retained as predictors because, despite their conceptual overlap, DIM captures the continuous lactation trajectory whilst SOL captures distinct stage-specific production patterns, providing complementary information to the model.

To ensure robust validation of the predictive models and maintain reproducibility of the results, the dataset was randomly split at the individual cow level using a fixed random state. To maintain good representation of individual cow data, 80% of records for each cow was used as the training set (364,862) and the remaining 20% as testing set (91,216). Two distinct subsets of predictor variables were generated for modelling purposes, enabling the assessment of the impact of environmental factors on the predictive performance of the models. The first subset incorporated weather-related variables, while the second subset excluded such factors. Data pre-processing involved treating categorical variables (breed, SOL, NL) as categorical features, transforming dates into datetime formats, data normalisation via standardisation of the numerical variables, and applying one-hot encoding to the categorical features. A machine learning pipeline was constructed to streamline the preprocessing and modelling phases.

Hyperparameter optimisation and grid search were conducted on the training data using both five-fold and ten-fold cross-validation strategies to optimise the predictive performance of each machine learning model. The scoring metric minimised during optimisation was the mean squared error (MSE), aiming to reduce the average squared difference between the predicted and actual MY. For both GBM (Friedman, 2001; Mason et al., 1999) and RF (Breiman, 2001; Liaw and Wiener, 2002) the hyperparameters adjusted were the number of estimators ranging from 100 to 500 in increments of 50, and maximum depth set at 5 and 10. The GBM model was tuned to learning rate at 0.01 and 0.05, while RF had fixed values for minimum samples required to split an internal node and at a leaf node, both set at 5. XGBoost (Chen and Guestrin, 2016) model was optimised by adjusting the number of estimators between ranges (100 to 500), the learning rate (0.01 and 0.05), and maximum depth (5 and 10), along with its two regularisation parameters, both set at 0.5.

The grid search systematically evaluated combinations of these hyperparameters across the different models and cross-validation strategies. The best model was identified based on the one with the lowest MSE and the optimal model parameters were found for each combination of model, subset, and cross-validation strategy. The overall best model was selected by comparing the best scores across all configurations. The selected best model was then trained using the optimal hyperparameters on the training data and evaluated on the testing data. Predictions were made on both training and testing sets and evaluation metrics, MSE, root mean squared error (RMSE) and R-squared (R^2), were used to assess the performance of each model.

2.4. Data sampling genetic algorithm

To select an appropriate dataset for testing the multi-objective optimisation algorithms, a genetic algorithm (GA) was employed to identify the optimal 30-day time window from the available data. This GA serves as a data selection tool, not a feeding optimisation algorithm. Its purpose is to find a period where a consistent set of cows in productive lactation stages are present throughout, providing a robust testing environment for the subsequent concentrate allocation optimisation. This 30-day window was chosen because it strikes a sensible balance: it is long enough for the model to adjust feed levels effectively, but still short enough to keep the optimisation computationally tractable (see Section 3.5). The dataset for optimisation was extracted from the preprocessed data with a focus placed on selecting the same set of cows that would be present throughout this period. Data were then filtered for cows within the DIM range of 60 to 270 days to target those in early to late lactation stages (critical periods for milk production optimisation).

The GA was configured with a population size of 50, a crossover rate of 0.8, a mutation rate of 0.07, an elite rate of 0.1, and was run over 50 generations. These parameter values were selected based on established practices for small combinatorial search problems. A population size of 50 was sufficient given the discrete and bounded nature of the search space, whilst a crossover rate of 0.8 and mutation rate of 0.07 reflect standard settings that balance exploration and exploitation for problems of this scale (Deb, 2001). Fifty generations were sufficient for this problem given the discrete and bounded nature of the search space, which consists of selecting start dates within a fixed time series. For simple combinatorial problems of this scale, 50 generations is a well-established starting configuration that provides adequate opportunity for the population to converge without unnecessary computational overhead. The window size was set to 30 days, corresponding to the duration for which consistent data was required. The fitness function was designed to prioritise windows that maximised the number of unique cows within the target DIM ranges (60–270 days), ensured cows were present on all days within the window, and minimised the variability of DIM values to achieve a balanced sample, the balance of DIM values measured by their standard deviation (with a lower standard deviation indicating a more uniform group), and the consistency of cow presence throughout the window. The fitness score was computed using a weighted sum of these factors, prioritising cows in the early DIM range and those consistently present, while penalising higher DIM variability to favour balanced groups.

Mathematical Formulation of the Fitness Function:

Let W represent a candidate 30-day window starting on day t. The fitness function maximises:

$$F(W) = w_1 \times N_{unique} - w_2 \times \sigma_{DIM} + w_3 \times C_{presence}.$$

where:

- N_{unique} = number of unique cows present on all 30 days within the target DIM range (60–270 days)
- σ_{DIM} = standard deviation of DIM values across selected cows (lower is better)
- $C_{presence}$ = consistency score (proportion of days each cow is present)
- w_1, w_2, w_3 = weight coefficients prioritising cow count, DIM balance, and presence consistency

The GA searches across possible starting dates t to find the window W^* that maximises F(W).

The GA proceeded through selection, crossover, and mutation processes. Selection involved choosing parents based on fitness scores, with a probability proportional to their fitness, ensuring that higher-performing chromosomes had a greater chance of contributing to the next generation. Crossover was performed using a single-point method, swapping start and end dates between parent chromosomes to create offspring, with validity checks to maintain the required window duration. Mutation introduced random adjustments to start dates within permissible ranges, maintaining the window size and introducing new genetic variations into the population. Elite chromosomes with the highest fitness scores were carried over unaltered to the next

generation, preserving the best solutions found so far and accelerating convergence.

2.5. Multi-objective optimisation for concentrate allocation

To maximise total MY through optimised concentrate allocation, a multi-objective optimisation approach was implemented and computed using the Pymoo library in Python (Blank and Deb, 2020). Four evolutionary algorithms, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) and the Strength Pareto Evolutionary Algorithm II (SPEA-II), the S-Metric Selection Evolutionary Multi-Objective Algorithm (SMS-EMOA), and the Reference Vector Guided Evolutionary Algorithm (RVEA), were utilised to solve the optimisation problem. The primary objective of this procedure was to determine the optimal daily grain-based concentrate allocation for individual cows to maximise total MY while minimising deviations in total concentrate usage. This dual-objective ensured that the optimisation not only focused on increasing milk production but also maintained the total concentrate allocation within specified constraint that ensured that the total concentrate allocated to all the cows did not significantly exceed or fall below the existing allocation levels, maintaining budgetary and nutritional balance among the cows.

Mathematical Formulation of the Optimisation Problem

The multi-objective optimisation problem is mathematically formulated as follows:

Decision Variables:

Let $x_{i,d}$ represent the daily concentrate allocation (kg/day) for cow i on day d, where:

$i = 1, 2, \ldots, N$ ($N = 165$ cows in our study).

$d = 1, 2, \ldots, D$ ($D = 30$ days).

Objectives:

1. Maximise total herd-level daily milk yield:

$$\text{Maximise } f_1 = \sum_{i=1}^{N} \widehat{MY}_{i,d}(x_{i,d}, F_{i,d})$$

where $\widehat{MY}_{i,d}$ is the predicted MY for cow i on day d, obtained from the pre-trained GBM model, and $\mathbf{F}_{i,d}$ represents the vector of other features (breed, DIM, SOL, NL, weather variables).

Average MY over the previous seven days was excluded from the optimisation feature vector because during optimisation these lagged values reflect historical yields generated under the original feeding regime rather than the hypothetical strategy being evaluated, introducing an inconsistency between predictor values and the concentrate allocations under consideration

2. Minimise deviation in total concentrate allocation:

$$\text{Minimise } f_2 = \left| \sum_{i=1}^{N} x_{i,d} - C_d^{\text{actual}} \right|$$

where C_d^{actual} is the actual total concentrate used on day d in the baseline data.

Constraints:

1. **Daily concentrate bounds per cow:**

$$L_{min} \le x_{i,d} \le L_{max}$$

where $L_{min} = 6$kg/day and $L_{max} = 11$kg/day

2. **Maximum daily change constraint:**

$$|x_{i,d} - x_{i,d-1}| \le 2 \text{ kg/day}$$

This ensures gradual daily adjustment within operationally feasible limits, reflecting widely accepted feeding management practices for Australian pasture-based dairy systems and maintaining continuity in individual cow feeding patterns.

3. **Total concentrate budget constraint:**

$$0.999 \times C_d^{\text{actual}} \le \sum_{i=1}^{N} x_{i,d} \le 1.000 \times C_d^{\text{actual}}$$

This maintains the overall concentrate usage within 99.9% to 100% of the baseline allocation.

4. **Adaptive bounds (for days $d > 1$):**

$$x_{i,d} \in \left[\max(L_{min}, x_{i,d-1} - 2), \min(L_{max}, x_{i,d-1} + 2)\right]$$

This combines the daily change constraint with the absolute bounds.

Problem Formulation:

$$\text{Maximise } f_1(x_{i,d})$$

$$\text{Minimise } f_2(x_{i,d})$$

$$\text{Subject to : Constraints } 1 - 4 \text{ above}$$

This formulation is solved using four evolutionary algorithms (NSGA-II, SPEA-II, SMS-EMOA, RVEA) to obtain a set of Pareto-optimal solutions representing different trade-offs between maximising MY and minimising concentrate deviation.

The dual formulation of concentrate allocation control, both as an optimisation objective (f_2) and as a hard constraint (99.9–100% of baseline), serves complementary purposes. The constraint establishes an absolute boundary ensuring that total concentrate usage remains within budgetary limits, preventing solutions that would require purchasing additional feed. Within this hard boundary, the objective function guides the evolutionary algorithms to explore the Pareto front, generating multiple solutions that represent different trade-offs between maximising milk yield (f_1) and maintaining concentrate allocation close to baseline levels (f_2). This approach allows farm managers to select from a range of Pareto-optimal solutions based on their risk tolerance and operational preferences. Solutions closer to 100% baseline allocation represent conservative strategies with minimal operational disruption, whilst those using the full allowable range (99.9–100%) maximise milk yield gains. The multi-objective framework thus provides flexibility that a constraint-only approach would not offer, as it generates a spectrum of viable feeding strategies rather than a single solution.

The constraints were also incorporated to reflect practical feeding limits on an individual cow level. Specifically, the concentrate for a cow on a given day could vary by a maximum of ± 2 kg/day from the concentrate allocation of the previous day, with overall minimum and maximum limits set at 6 kg/day and 11 kg/day, respectively. These limits were based on widely accepted feeding management practices for Australian pasture-based dairy systems ("Feeding dairy cows : a manual for use in the Target 10 Nutrition Program," 2015). Additionally, the sum of individual concentrate allocations was constrained to be within 99.9% to 100% of the total actual concentrate used in the baseline data for that day, ensuring that the overall concentrate usage remained consistent with existing feeding regime. The tight herd-level constraint (99.9–100%) reflects conservative management practice ensuring that optimisation improvements are achieved through reallocation rather than requiring additional feed purchases, thereby maintaining operational feasibility and budget adherence.

The optimisation process involved several key steps. Data preparation was undertaken using the daily records of cows over the previously identified optimal 30-day window from the GA. Features included in the optimisation were breed, DIM, SOL, NL, `vp`, min_temp, max_temp, rh_tmin, rh_tmax, and CONC. The pre-trained predictive model for MY, developed earlier using a Gradient Boosting Regressor, was loaded

to predict MY based on these predictor features and the proposed concentrate allocations.

A custom optimisation problem class was defined, encapsulating the objectives and constraints of the optimisation. This class utilised the Pymoo framework because of its capabilities to handle multi-objective problems with constraints. All the algorithms were configured with a population size of 200. Genetic operators included a simulated binary crossover (SBX) with a crossover probability of 0.95 and an eta value of 15, and polynomial mutation with an eta value of 20. Duplicate solutions were eliminated to maintain population diversity, and the termination criterion was set to 100 generations to balance computational efficiency and solution quality. Infeasible solutions generated during crossover or mutation, those violating the daily concentrate bounds or the total budget constraint, were handled automatically by the Pymoo framework through the constraint violation handling mechanism defined in the problem class, which penalised infeasible candidates during selection and guided the search towards feasible regions of the solution space. All model training and optimisation runs were executed on a personal computer equipped with a 13th Generation Intel Core i9-13900HX processor (2.20 GHz, 24 cores, 32 logical processors), 32.0 GB RAM, running Microsoft Windows 11 Home.

To ensure statistical robustness and account for the stochastic nature of evolutionary algorithms, each algorithm (NSGA-II, SPEA-II, SMS-EMOA, RVEA) was executed 10 times using different random seeds (3, 7, 15, 22, 28, 31, 38, 41, 46, 49). This approach enabled statistical validation of results through normality tests, analysis of variance (ANOVA), and post-hoc pairwise comparisons. The optimisation was executed iteratively for each day within the 30-day window. For the first day, the initial population or guesses was seeded with the actual concentrate allocations from the sample data. For subsequent days, the initial solutions were based on the optimised allocations from the previous day, adhering to the dynamic bounds. This approach was adopted to aid continuity and practicality in the daily feeding adjustments. In each iteration, candidate solutions, sets of individual cow concentrate allocations, were evaluated using the predictive model to estimate the resulting MY. The objectives and constraints were computed accordingly, and the optimisation algorithms guided the search towards Pareto-optimal solutions that balanced MY maximisation and concentrate allocation minimisation.

Upon completion of the optimisation process for each day, a single solution was selected from the Pareto front of non-dominated solutions. The selection criterion prioritised solutions that maximised MY (objective 1) while maintaining minimal deviation from baseline concentrate allocation (objective 2), effectively choosing the solution closest to the ideal point where both objectives are optimised simultaneously while satisfying all constraints. The optimised concentrate allocations and the predicted MY were recorded. Key performance metrics for each optimisation algorithm (NSGA-II, SMS-EMOA, RVEA and SPEA-II), including the percentage increase in predicted MY compared to the actual sample data, the total optimised concentrate used, and the optimisation compute time, were calculated for each day and used for comparative analysis of their performance. Fig. 1 provides an overview of the complete methodological framework, illustrating the four stages, their outcomes, and how they connect. This clarifies the role of the Data Sampling GA (Stage 3) as a data selection tool that identifies the optimal testing window for the multi-objective optimisation algorithms (Stage 4), while the pre-trained prediction model from Stage 2 serves as the objective function for optimisation.

3. Results and discussion

3.1. Descriptive analysis

After the raw dataset was cleaned and pre-processed, the dataset was comprised of 456,078 records, detailing milk production and environmental factors for 1053 unique dairy cows across 885 sampling days. It included a broad range of records per cow, reflecting varied DIM, CONC, and MY records. The DIM values ranged from 1 to 516 days, with an average of 180.61 ± 109.20 days, capturing cows at different SOL, from early (<100 DIM), mid (100 to 199 DIM) to late (200 to 305 DIM) and also extended lactation (>305 DIM). This comprehensive dispersion in DIM was critical for modelling and optimising MY, as lactation stage significantly impacts milk production. The concentrate allocation showed a wide variability with values ranging from 2 to 14.5 kg/cow/day with an average of 6.98 ± 2.24 kg/cow/day. This variability in concentrate distribution was beneficial for the MY modelling and optimisation process as it allowed for analysis of how different feeding levels influenced MY across various stages of lactation and breeds. Given the variability in MY among cows and the fact that cows were grazed on pasture, MY per unit of concentrate fed presented notable variation, ranging from 0.17 to 15.80 L/kg of concentrate/cow/day, with an average of 2.64 ± 1.19 L/kg of concentrate/cow/day.

The analysis of categorical data revealed insightful distributions across breed, NL, and SOL. The majority of records were from Holstein cows, comprising 39.62%, and this was followed by Holstein Friesian at 31.30%. The Friesian x Jersey crossbreed accounted for 16.06%, while the Swedish Red-and-White represented 13.02%. This breed composition indicates a predominant focus on Holstein and related high-yielding breeds in the dataset. Regarding the distribution by NL, first-lactation cows constituted the largest group, contributing 34.33% of the total, signifying a significant representation of younger cows followed by second-lactation (26.60%), fourth-lactation (22.65%), and third-lactation (16.43%). For the SOL, late (including extended) lactation cows were most prevalent, comprising 41.85% of the dataset, with mid lactation and early lactation cows contributing 30.00% and 28.14%, respectively. This distribution provides a well-rounded dataset for examining productivity across multiple lactation cycles and lactation stages.

3.2. Correlation analysis

The correlation analysis revealed that several variables were strongly correlated with MY. As expected, average MY over the previous seven days (0.81) exhibited the strongest positive correlations with MY, indicating that historical MY metrics are closely aligned with current milk output (Fig. 2). Concentrate consumed per cow also had a moderate positive correlation (0.35) with MY, showing the impact of concentrate intake on milk production levels. Conversely, DIM showed a significant negative correlation with MY (−0.45), reflecting the expected decline in MY as lactation progresses. Other environmental factors, such as `vp`, min_temp and max_temp demonstrated weak correlations with MY, implying that while they may affect cow welfare, their direct impact on daily MY is minimal. All correlations presented in Fig. 2 were statistically significant ($p < 0.05$). Overall, this correlation analysis highlights that historical milk performance metrics, feeding levels, and NL are key driving factors correlated with MY, while environmental conditions play a more limited role. Additionally, visualising the relationship between average MY per kg of concentrate and DIM across different NL and breeds, as shown in Fig. 3, the relationship indicated that the average MY per kg of concentrate consumed ranged between 1 and 6 L/kg of CONC showing that there is still a wide variability in feed efficiency among the cows.

3.3. K-means clustering

The K-means clustering analysis, performed separately for each NL, revealed distinct production heterogeneity within each parity group. The optimal cluster number was determined to be k = 2 through comprehensive mathematical validation using three complementary metrics: the elbow method, silhouette scores (ranging from 0.494 to 0.523 across lactation numbers), and Davies-Bouldin Index (all values <0.76), with full validation results presented in Supplementary Fig. S1. This k = 2 configuration consistently outperformed higher

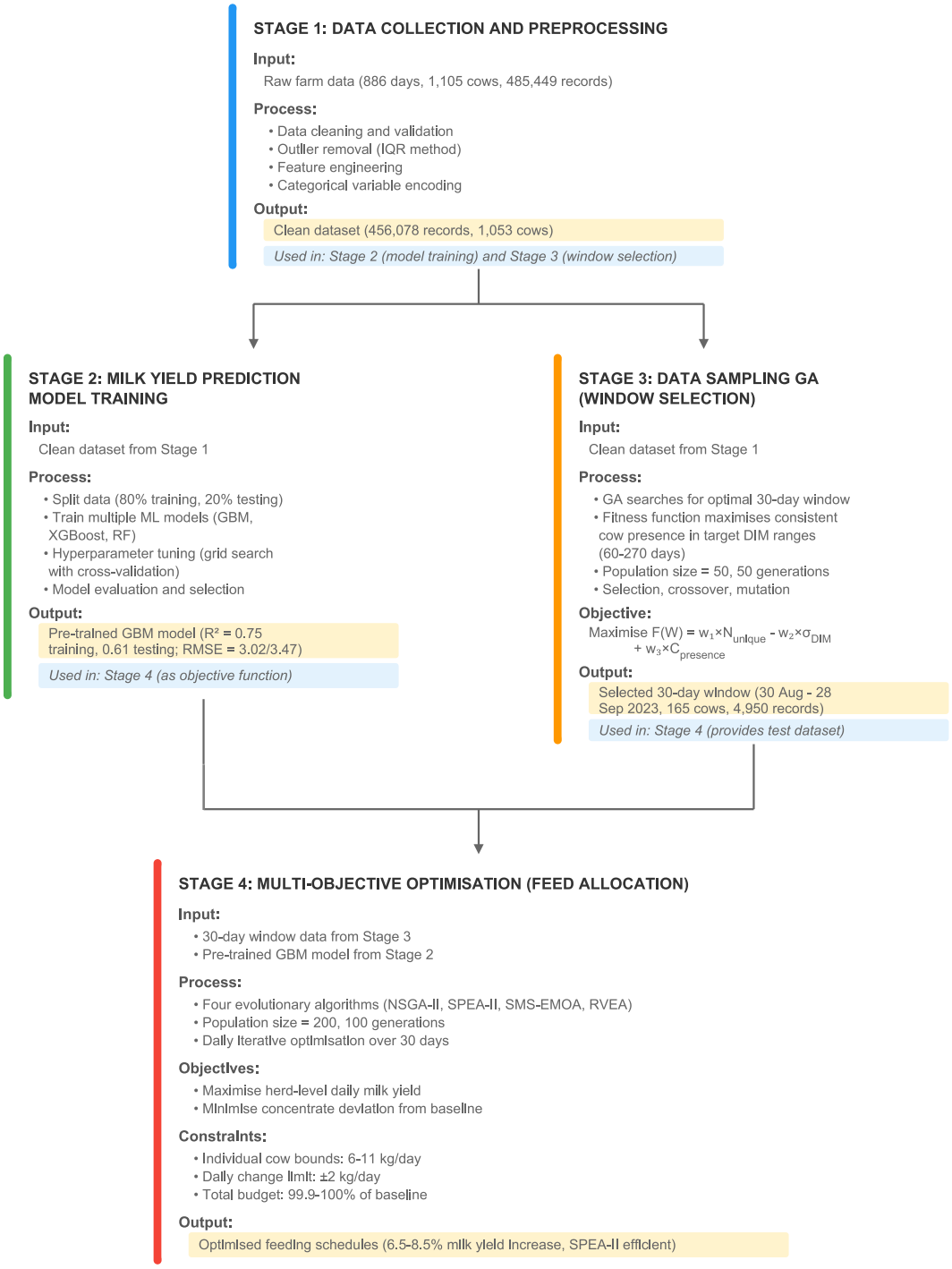


Fig. 1. Methodological framework showing the four-stage workflow. Stage 1: Data preprocessing; Stage 2: Machine learning model training (GBM selected as best model); Stage 3: Data Sampling GA for optimal 30-day window selection (165 cows); Stage 4: Multi-objective optimisation using evolutionary algorithms integrated with the pre-trained model. Arrows indicate data flow between stages, demonstrating how the Data Sampling GA (Stage 3) serves as a data selection tool for testing the optimisation algorithms (Stage 4).

cluster numbers across all validation criteria, indicating that cows within each parity naturally separate into two distinct production groups based on their average MY and concentrate intake patterns.

Within each NL, the two clusters represented cows with relatively higher production levels (higher MY and higher CONC) and cows with relatively lower production levels. For first-lactation cows (NL = 1), the higher-producing group ($n = 430$) averaged 15.3 L/cow/day with 7.1 kg/cow/day of concentrate, whilst the lower-producing group ($n = 237$) averaged 11.4 L/cow/day with 5.9 kg/cow/day of concentrate (Supplementary Table S1). This production differential was even more pronounced in multiparous cows. Second-lactation cows (NL = 2) showed the highest production in the higher-performing cluster (18.8 L/cow/day with 7.4 kg/cow/day concentrate, $n = 371$) compared to their lower-performing counterparts (13.8 L/cow/day with 6.3 kg/cow/day concentrate, $n = 157$). Third- and fourth-lactation cows exhibited similar patterns, with higher-producing clusters achieving 20.3 and 19.7 L/cow/day respectively, consistently demonstrating that multiparous cows in the higher-production groups benefitted most from increased concentrate allocation and were potentially more productive overall.

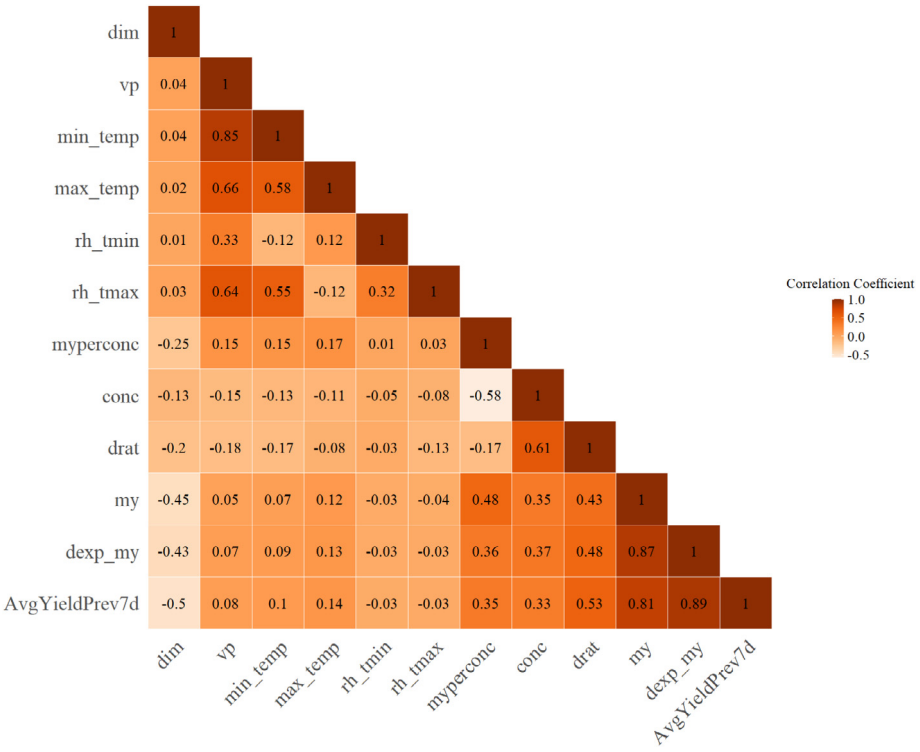


Fig. 2. Correlation heatmap of key variables influencing milk yield (MY). A heatmap illustrating the correlation matrix for key production and environmental variables associated with MY. The strongest positive correlations are observed between MY and historical milk yield indicators (AvgYieldPrev7d), while DIM shows a significant negative correlation, reflecting the decline in MY over the lactation period and moderate positive correlations with concentrate intake (CONC). Environmental factors(vp, min_temp, max_temp, rh_tmin, rh_tmax) exhibited weaker correlations, suggesting a more indirect influence on daily milk production.

The clustering analysis revealed that multiparous cows generally achieved higher average MY, particularly within the higher-producing clusters, consistent with the multiparous profiles described by Lee et al. (2020); who reported steeper lactation-curve peaks and greater total yields once cows moved beyond their first lactation. In contrast; first-lactation cows displayed lower overall production levels with the lower-producing cluster averaging just 11.4 L/cow/day; aligning with the flatter; more persistent lactation curves reported for primiparous cows by Lee et al. (2020) who used k-medoids clustering and found that primiparous cows exhibited flatter curves with lower peak yields and later peak timing, whilst multiparous cows showed steeper peaks and higher production. Our k = 2 results directly support this: first-lactation cows (11.4–15.3 L/cow/day) produced substantially less than multiparous cows (18.8–20.3 L/cow/day in higher-producing clusters), demonstrating the lower production capacity and flatter profile characteristic of primiparous animals. The presence of substantial within-cluster variability, as indicated by the within-cluster sum of squares values (withinss) ranging from 887.25 to 1948.89 across the eight cluster combinations (Supplementary Table S1), suggests considerable individual variation even within production groups. This heterogeneity highlights opportunities for further individualised feeding optimisation strategies to standardise and enhance productivity across different production levels.

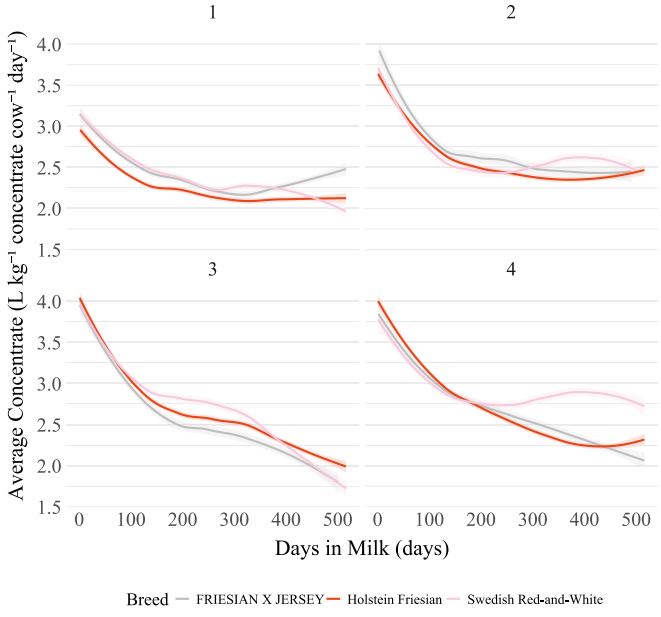


Fig. 3. Average milk yield per kg of concentrate by days in milk (DIM) across different breeds and number of lactations (NL). The graph illustrates the relationship between the average milk yield per kg of concentrate consumed (L/kg concentrate cow^{-1} day^{-1}) and DIM for different breeds (smooth trend lines with shaded ribbons showing the 95% confidence interval), with each subplot representing a specific NL.

The k = 2 clustering effectively captured the fundamental production dichotomy within each parity: cows respond differently to concentrate supplementation based on their inherent production capacity. As shown in Fig. 4, the two clusters exhibited distinct separation patterns across all lactation numbers, with the clusters representing high and low producers within each NL rather than different stages of lactation. This production-based segmentation, validated through robust mathematical metrics, provides a practical framework for precision feeding strategies. By identifying cows with similar production characteristics within each parity, the clustering facilitates targeted nutritional management, whereby higher-producing cows receiving optimal concentrate levels to support their elevated milk output, whilst lower-producing cows receive appropriately calibrated supplementation, ultimately improving overall feed use efficiency and resource allocation across the herd.

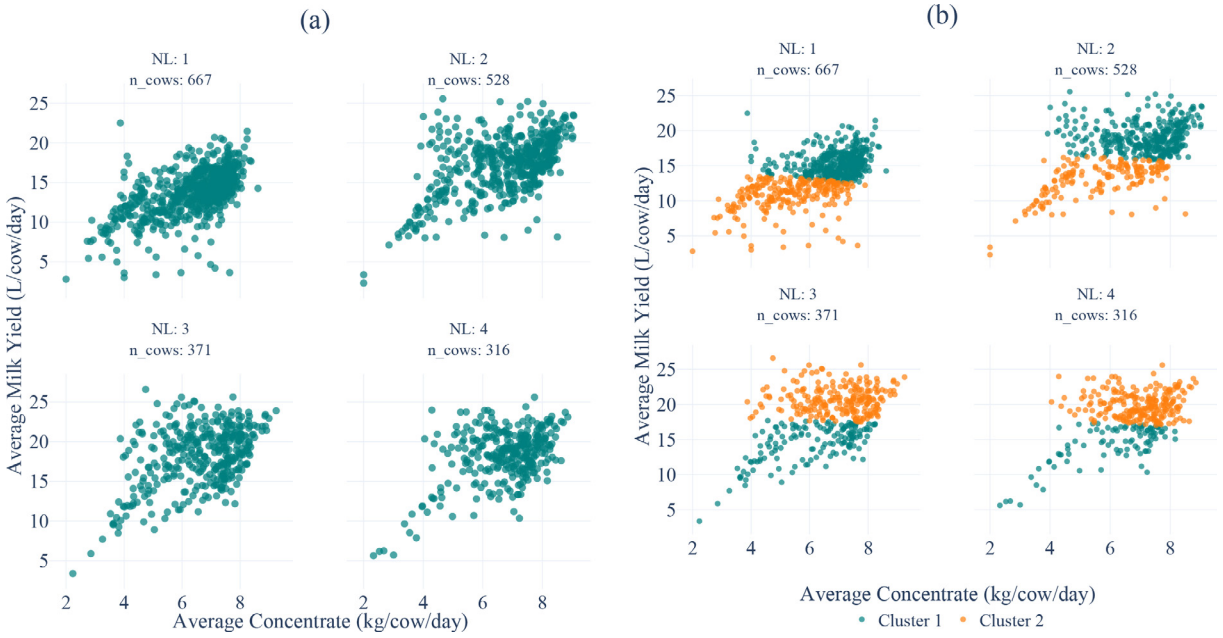


Fig. 4. Relationship between MY and concentrate intake by number of lactations (NL) with K-Means Clustering. Plot (a) shows the distribution of individual cows before clustering, while plot (b) demonstrates the results of K-Means clustering using k=2 clusters, highlighting distinct production patterns between high- and low-producing groups within each lactation number. The optimal cluster number (k = 2) was determined through mathematical validation using the elbow method, silhouette score, and Davies-Bouldin Index, performed separately for each NL (validation results shown in Supplementary Fig. S1). Cluster 1 (teal) and Cluster 2 (orange) represent two distinct production groups within each parity, with detailed cluster statistics provided in Supplementary Table S1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Performance evaluation of the MY prediction model

The modelling results after training demonstrated the predictive performance of various machine learning models in estimating MY with two feature subsets: *'w'* (with cow characteristics and environmental data) and *'nw'* (with only cow characteristics). The GBM, applied to the full set of predictors (cow characteristics and environmental data) with 10-fold cross-validation, emerged as the best-performing model. It achieved the lowest MSE of 12.26, an RMSE of approximately 3.5 L/cow/day, and an R^2 of 0.75 (Fig. 5a). These metrics indicated precise predictive capability, high reliability, and ability of the GBM in explaining the variance in MY. The optimised hyperparameters for the

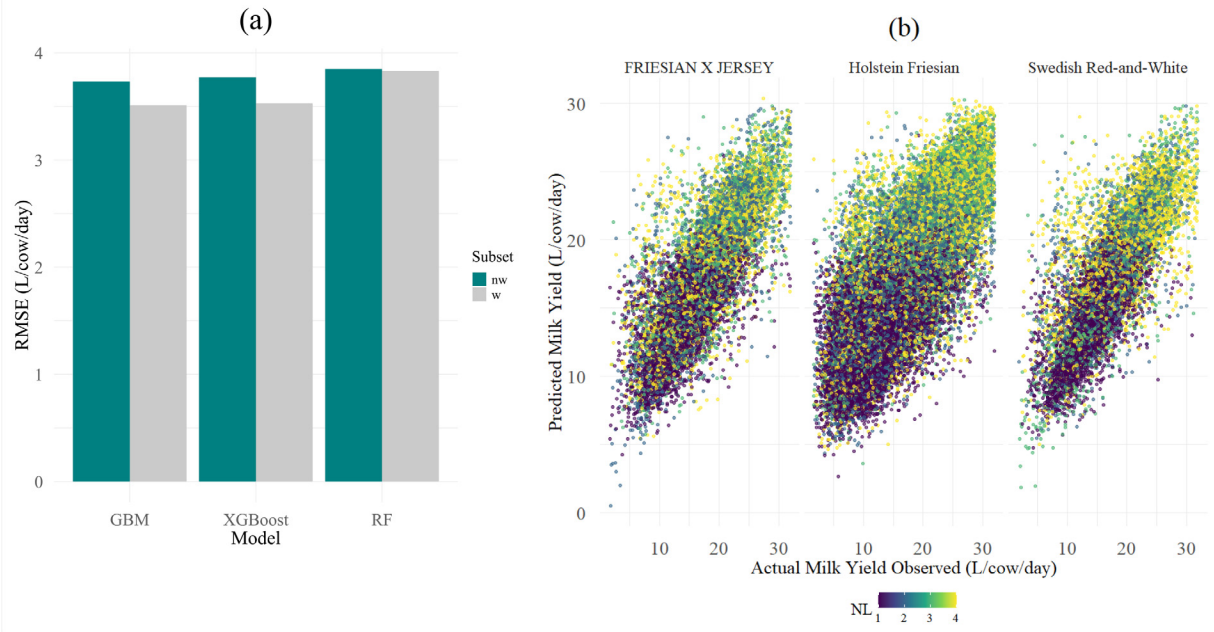


Fig. 5. Comparative analysis of model performance metrics and MY predictions. (a) Bar plot displaying the Root Mean Squared Error (RMSE) of different machine learning models (GBM, XGBoost, and RF) for predicting milk yield for the trained model. The RMSE, expressed in litres per cow per day (L/cow/day), is the square root of the mean squared error and quantifies the average prediction error. Results are presented for models trained on two feature subsets: *'w'* (with cow characteristics and weather data) and *'nw'* (with only cow characteristics). (b) Scatter plots depicting predicted versus observed milk yield across various cow breeds, with NL represented as a gradient colour scale. These plots illustrate the predictive accuracy of the best-performing model and how it performs on the test set across different breeds and NL.

GBM included a learning rate of 0.05, a maximum depth of 10, a minimum of 5 samples per leaf, 5 samples per split, and 500 estimators. This fine-tuning contributed significantly to its performance, providing a balance between model complexity and generalisability. The XGBoost model also exhibited competitive performance when applied to the full set of predictors, recording an MSE of 12.39 and an RMSE only marginally higher than that of the GBM. In contrast, the RF model showed higher MSE and RMSE values, with an MSE of approximately 14.67, an RMSE of around 3.83 L/cow/day, and an R^2 of 0.68 across both feature subsets. This indicates that while RF maintained reasonable predictive accuracy, it was outperformed by the boosting algorithms.

After training, the GBM model was evaluated on the test set, achieving an R^2 of 0.61 (explaining 61% of variance in MY), MSE of 12.05, RMSE of 3.47 L/cow/day, and a Concordance Correlation Coefficient (CCC) of 0.75 (0.81 for training). The CCC, which assesses both precision and accuracy, indicates substantial agreement between predicted and observed values. These performance metrics demonstrate the effectiveness of the model in predicting MY on unseen data and are comparable to similar studies where Streefland et al. (2023) achieved $R^2 = 0.71$; and Ji et al. (2022) reported RMSE of 3.2–4.5 L/cow/day using ML for MY prediction.

Feature importance analysis from the best-performing GBM model revealed the relative contribution of predictor variables to milk yield prediction. Days in milk was the most important predictor (relative importance: 42.4%), followed closely by concentrate intake (CONC, 37.1%). Environmental variables collectively contributed 18.7% to predictive accuracy, with maximum temperature (5.8%) and minimum temperature (4.9%) being the most influential climatic factors. Breed contributed 1.7%, whilst stage of lactation accounted for a negligible proportion (0.1%) of model predictions. These importance values reflect the relative influence of each variable on the model decision-making process, with DIM serving as the strongest predictor whilst concentrate intake represents the primary modifiable factor available for optimisation, accounting for more than one-third of the predictive power of the model.

The modelling analysis provided key insights into MY predictions across different SOL, highlighting how specific features influence predictive performance. As shown in Table 1, average predicted and actual MY consistently rise with increasing parity, a trend that is also evident in Fig. 5b, where the two clusters (low and high producers) show distinct production patterns within each lactation number. An additional 2 kg of concentrate delivers the greatest marginal return in second-lactation cows, lifting predicted yield by about 2.5 L/cow/day, whereas the same supplement yields smaller gains in other parities. Temperature effects are similarly parity-dependent: maximum temperatures above 26 °C and minimum temperatures exceeding 19 °C raise predictions by roughly 1–3 L/day, but only once cows have moved beyond their first lactation. Second-parity cows showed substantial productivity, with mid-lactation animals achieving mean predicted yields of 18.4 L/day at an average concentrate allocation of 7.7 kg/day (Supplementary Table S2). These animals exhibited a mean milk-to-concentrate conversion efficiency of 2.60 L/kg, intermediate between the higher efficiency observed in early lactation (3.11 L/kg) and the lower efficiency in late lactation (2.44 L/kg). Across all parity, the early-lactation phase remains the most responsive period, adding 4 to 5 L per cow per day above baseline and highlighting the value of targeted nutritional support during the first three months after calving. Table 1 shows predicted MY values and the marginal effects of concentrate supplementation and temperature on daily milk production across different parities (NL) and SOL.

Building on these results, our analysis demonstrates that models leveraging boosting techniques and ensembles, particularly the GBM fine-tuned with optimal hyperparameters, provide superior predictive accuracy and lower error margins. This level of performance is crucial for reliable MY forecasting in dairy production, supporting effective decision-making and resource optimisation and supported by previous research done by (Streefland et al., 2023). The results of the correlation analysis and predictive modelling align with the findings of previous research conducted on the impact of environmental and cow-specific factors on MY (Azubuike et al., 2024). In particular; the weak correlations observed between MY and certain environmental variables; such as vapour pressure; temperature; suggest that while these factors may influence cow comfort and welfare; their direct effect on daily MY is limited (Fig. 2). This is consistent with the work by Zhang et al. (2020), which found that the introduction of weather parameters had a relatively small impact on the accuracy of MY forecasting, with only minor improvements observed.

Furthermore, Gasser et al. (2023) reported that, in moderate climates, the inclusion of meteorological variables such as temperature, rainfall, wind speed, and barometric pressure did not improve MY prediction models. Their study indicated that models incorporating lagged MY records were more effective in predicting future MY, suggesting that meteorological data may be indirectly captured through lagged MY production. This aligns with our results, where the inclusion of weather parameters did not result in huge improvements in model performance, as evidenced by the minimal decrease in MSE and RMSE values for the models with weather data compared to those without (Fig. 5a) the weather data.

However, the predictive models evaluated in our study still demonstrated the potential of incorporating both cow characteristics and environmental or weather data to effectively predict MY. This is consistent with the findings by Katsini et al. (2024); who emphasised the role of climate factors such as humidity; wind; and radiation in influencing milk production. The comparison of the models (GBM; XGBoost; and RF) with and without environmental data (w vs. nw) supports the notion that while including environmental factors can enhance the predictive accuracy of the models; the magnitude of this improvement remains limited (Fig. 5a); reflecting similar conclusions drawn by previous studies regarding the effectiveness of weather parameters in forecasting MY (Gasser et al., 2023).

Table 1
Model-predicted milk yield and key covariate effects across parities and lactation stages.

Lactation (NL)	Lactation stage	DIM range (days)	Mean predicted MY (L cow^{-1} d^{-1})	Typical increase over baseline* (L cow^{-1} d^{-1})	Effect (Δ MY) of an extra ≈ 2 kg concentrate	Temperature-related effect
1	Early	12–95	13.76	2.41	+1.97 L	Tmax 26.3–35.5°C → +1 L
1	Mid	95–207	≈14.7†	1	NS	NS
2	Early	12–95	16.31	4.42	+2.53 L	Tmax >26.5°C → +1.2 L
2	Mid	98–214	18.54	+0.9 (vs overall 17.64)	NS	NS
3	Early	12–95	16.85	5.3	NS	Tmin >19.3°C → +2.63 L; Tmax 26.5–36.2 °C → +1.89 L
4	Early	12–95	17.25	4.96	NS	NS

NS = No significant effect detected or effect size below reporting threshold. T_max = daily maximum ambient temperature; T_min = daily minimum ambient temperature.

* Baseline represents the predicted milk yield at reference conditions: early-mid lactation (DIM 50–150), moderate concentrate levels (~6 kg/day), and ambient temperatures (~20 °C). The "Typical increase" indicates additional milk yield when favourable conditions are present (e.g., higher concentrate, optimal temperatures) compared to baseline.

† Derived as early-lactation mean + 1 L attributed to mid-lactation stage effect.

Additionally, the results indicate that cow characteristics can be used to predict MY effectively, challenging the conventional assumption that liveweight, a factor traditionally regarded as key for predicting MY and feed intake, is indispensable for accurate model training. The performance of the models, particularly GBM and XGBoost, in the absence of liveweight data demonstrates that other cow traits and environmental variables can offer a reasonably reliable basis for MY prediction. Our model achieved an R^2 of 0.61 and RMSE of 3.47 L/cow·day^{-1} on the test set, which is comparable to models reported in similar studies. While the inclusion of liveweight data might improve the predictive process, the models appear to function adequately without it (Berry et al., 2007). This is a practical consideration and especially noteworthy, given that liveweight data can be difficult to obtain consistently on many conventional farms, particularly in pasture-based systems where cows are not routinely handled. The logistical burden and cost of frequent liveweight measurements make our approach more feasible for real-world implementation. Ultimately, these findings suggest that although access to liveweight data could enhance MY forecasts, the broader applicability of data-driven approaches in dairy management for the future remains feasible with access to larger datasets even when relying on fewer, yet strategically important, variables such as DIM (42.4% relative importance) and concentrate intake (37.1%), which together capture the physiological trajectory and nutritional state of each cow and can adequately represent liveweight as proxy variables in data-rich commercial datasets.

3.5. Evolutionary algorithms for supplement optimisation

The genetic algorithm used for sample selection resulted in a dataset comprising 4950 records over the 30-day window. After 50 generations, the best solution identified by the GA was a 30-day window from 30 August to 28 September 2023. This period maximised the number of unique cows within the desired DIM ranges and ensured a substantial number of cows were present on all days. Out of the whole herd, 165 cows were consistently present throughout this period and were selected as a good sample for the optimisation of grain-based concentrate allocation satisfying the stipulated constraints (Fig. 6). The statistical summary of the sampled data is shown in Table 2, which details key production factors including average MY, concentrate intake, and environmental variables. This summary highlights the representativeness and distribution of the sample obtained from the GA to aid accurate modelling and analysis for concentrate optimisation among the cows selected.

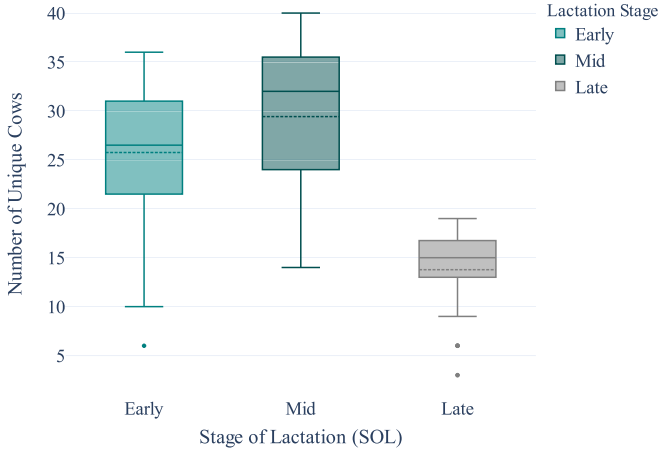


Fig. 6. Distribution of unique cows across Days in Milk (DIM) Range (60–270 Days) in the sampled data for the optimisation process. The box plots illustrate the distribution of unique cows ($n = 165$) by their stage of lactation within the range of 60 to 270 DIM, based on sampled data selected for the optimisation process. The data represents the optimal 30-day window identified from 30 August 2023 to 28 September 2023, derived using a genetic algorithm optimisation.

Table 2
Statistical summary of the sampled data used for optimisation analysis in this study.

Variables	Min	Max	Mean	Median	Std.	IQR
Days in Milk (days)	60	270	149.18	144	52.77	76
Concentrate intake (kg/cow/day)	3.87	13.74	8.86	9.71	1.97	3.2
Milk Yield (L/cow/day)	4.09	31.98	20.26	20.43	4.74	6.48
Average MY for previous 7 days (L/cow)	8.82	30.45	20.46	20.51	3.85	5.67
Milk Yield per Concentrate (L/kg/cow/day)	0.5	8.17	2.42	2.21	0.9	0.79
Vapour pressure (kPa)	5.4	16.9	10.51	10.9	2.67	4.2
Minimum temperature (°C)	−1.9	16.9	5.77	5.3	4.15	6.3
Maximum temperature (°C)	17.5	36	23.85	22.25	5.02	6
Minimum relative humidity (%)	37.9	100	91.27	100	18.26	0
Maximum relative humidity (%)	13.1	75.5	36.77	34.5	12.69	18.2

The table provides a statistical summary of the key variables, covering cow characteristics and related environmental variables, for the 165 cows selected by the GA for the optimisation process. The metrics shown include minimum (Min), maximum (Max), mean, median, standard deviation (Std.), and interquartile range (IQR).

To ensure statistical robustness of algorithm comparisons, each evolutionary algorithm was executed 10 times using different random seeds (3, 7, 15, 22, 28, 31, 38, 41, 46, 49) (Fig. 7). Shapiro-Wilk normality tests confirmed that milk yield increase distributions were normal for all algorithms ($p > 0.05$), enabling parametric statistical analysis. One-way ANOVA revealed statistically significant differences among the four algorithms ($F = 303.24$, $p < 0.001$). Tukey HSD post-hoc pairwise comparisons demonstrated that all algorithms differed significantly from each other in milk yield improvement (all $p < 0.001$). NSGA-II achieved the highest mean milk yield increase of 8.64 ± 0.10% (range: 8.43–8.75%), significantly outperforming SPEA-II (7.94 ± 0.13%, range: 7.73–8.13%), RVEA (7.41 ± 0.10%, range: 7.26–7.56%), and SMS-EMOA (6.63 ± 0.24%, range: 6.31–7.13%) (Fig. 7). The low standard deviations indicate consistent performance across runs. For computational efficiency, optimisation time distributions were non-normal (Shapiro-Wilk test, $p < 0.05$ for SPEA-II and RVEA), necessitating Kruskal-Wallis non-parametric testing. Results showed NSGA-II was significantly faster (74.9 ± 0.7 s per day) compared to SPEA-II (199.5 ± 51.0 s/day, $p = 0.001$), RVEA (205.8 ± 14.7 s/day, $p = 0.022$), and SMS-EMOA (220.7 ± 8.1 s/day, $p < 0.001$). No significant differences in computation time existed among SPEA-II, RVEA, and SMS-EMOA (all $p > 0.05$). Complete statistical results including descriptive statistics, normality tests, and post-hoc comparisons are provided in Supplementary Table S3.

The baseline feeding regime implemented through the Delpro management system represents a traditional response-based feeding approach commonly used in commercial dairy operations (Dale et al., 2016; Henriksen et al., 2019). By comparing our evolutionary algorithm optimisation against this established system, we provide a fair benchmark of improvement over current industry practices. The 6.6–8.8% increase in MY achieved by our approach demonstrates potential gains in feed conversion efficiency compared to conventional decision support systems.

The results are consistent with the literature on multi-objective optimisation in dairy systems, particularly with regard to the SPEA II algorithm. Notte et al. (2021) similarly found that although SPEA II provided high-quality feed resource allocation results; it was computationally inefficient for large-scale applications. In contrast; the integration of pre-trained machine learning models with evolutionary algorithms in this study led to notable improvements in computational performance. Among the four algorithms tested; NSGA-II demonstrated the best

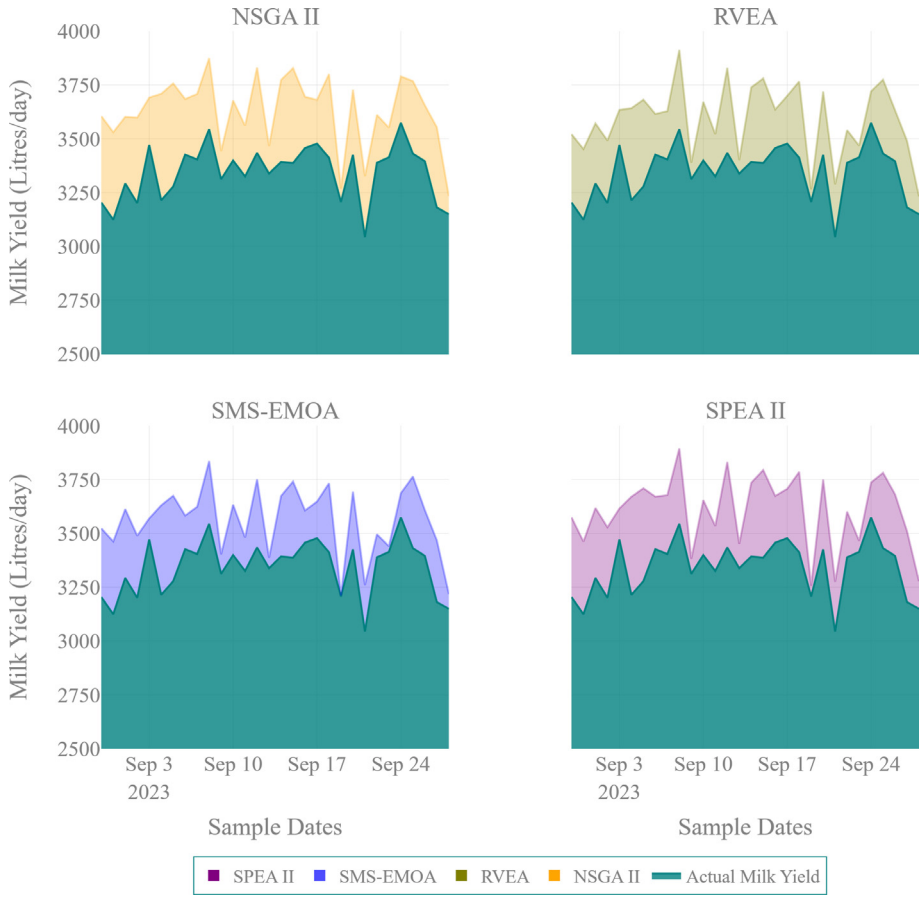


Fig. 7. Milk yield comparison and optimisation across evolutionary algorithms over a 30-day period for 165 cows.
Fig. 7 presents a comparison of the actual and predicted milk yield for the four evolutionary algorithms used in the optimisation process: Non-dominated Sorting Genetic Algorithm II (NSGA II), Reference Vector Guided Evolutionary Algorithm (RVEA), S-Metric Selection Evolutionary Multi-Objective Algorithm (SMS-EMOA), and Strength Pareto Evolutionary Algorithm II (SPEA II). The actual milk yield is plotted in teal, while the predicted milk yield is shown for each algorithm in distinct colours—orange for NSGA II, olive-green for RVEA, blue for SMS-EMOA, and purple for SPEA II. The area charts depict the trend of both actual and predicted milk yields over time, with the area under the curves providing a visual representation of the yield variations. This figure illustrates the effectiveness of each algorithm in predicting milk yield, enabling a direct comparison of their performance and optimisation results. Results shown are from one representative run (random seed = 3) out of 10 independent runs performed for statistical validation (see Fig. 7 and Supplementary Table S3 for complete statistical analysis across all runs). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

balance of prediction accuracy and computational efficiency; while SPEA-II also showed substantial improvements over traditional optimisation approaches (Fig. 8); addressing the computational limitations highlighted in previous research. This approach accelerated the optimisation process while maintaining the accuracy of MY predictions; thereby optimising the trade-off between speed; performance and the accuracy of MY predictions. Such an advancement represents a significant step in overcoming the challenges associated with slower optimisation times; as highlighted in previous research; such as that by Notte et al. (2021) who noted the inefficiency of SPEA II for large-scale dairy system applications.

Furthermore, this integration of machine learning into the evolutionary optimisation framework in this study provides a promising direction for future research in dairy farm optimisation. The combination of evolutionary algorithms and machine learning, as seen in the work of (Das et al., 2023); suggests that adaptive and self-learning techniques can improve performance in complex dairy feed optimisation problems. Our findings support this approach; highlighting the scalability and adaptability of such methods for dairy systems. As Das et al. (2023) demonstrated with their SMP-JaQA algorithm, optimising feed with a balance between computational efficiency and accuracy remains a core challenge. Our integration of machine learning reflects this trend, enhancing the overall robustness and efficiency of the optimisation process.

In comparison to Campos et al. (2023); they demonstrated that algorithms tailored to the individual characteristics of cows; such as optimising feeding for each cow; result in better milk production outcomes. This aligns with our approach; where individual cow characteristics were considered to optimise concentrate allocation and; consequently; MY. The success of their approach in optimising individual cow diets further supports the importance of personalised strategies in maximising MY; particularly when considering the variability of cow responses to feeding strategies. Souza et al. (2022) reported similar findings, emphasising the necessity of personalised feeding recommendations to improve feed efficiency and overall milk production. This reinforces the notion that algorithms accounting for individual cow data can optimise feeding and increase MY.

Overall, this study successfully integrated machine learning models with evolutionary algorithms to optimise grain-based concentrate allocation by incorporating individual cow data, feeding patterns, and environmental factors. Additionally, data-driven approaches combining ML with evolutionary optimisation can effectively address major dairy farming challenges, ultimately enhancing MY and feed use efficiency. While formal economic analysis is beyond the scope of this study, the 6.6–8.6%, increase in MY without additional concentrate expenditure suggests potential for improved farm profitability, though actual economic benefits would depend on factors such as milk prices, feed costs, and implementation expenses. The analysis confirmed that ML

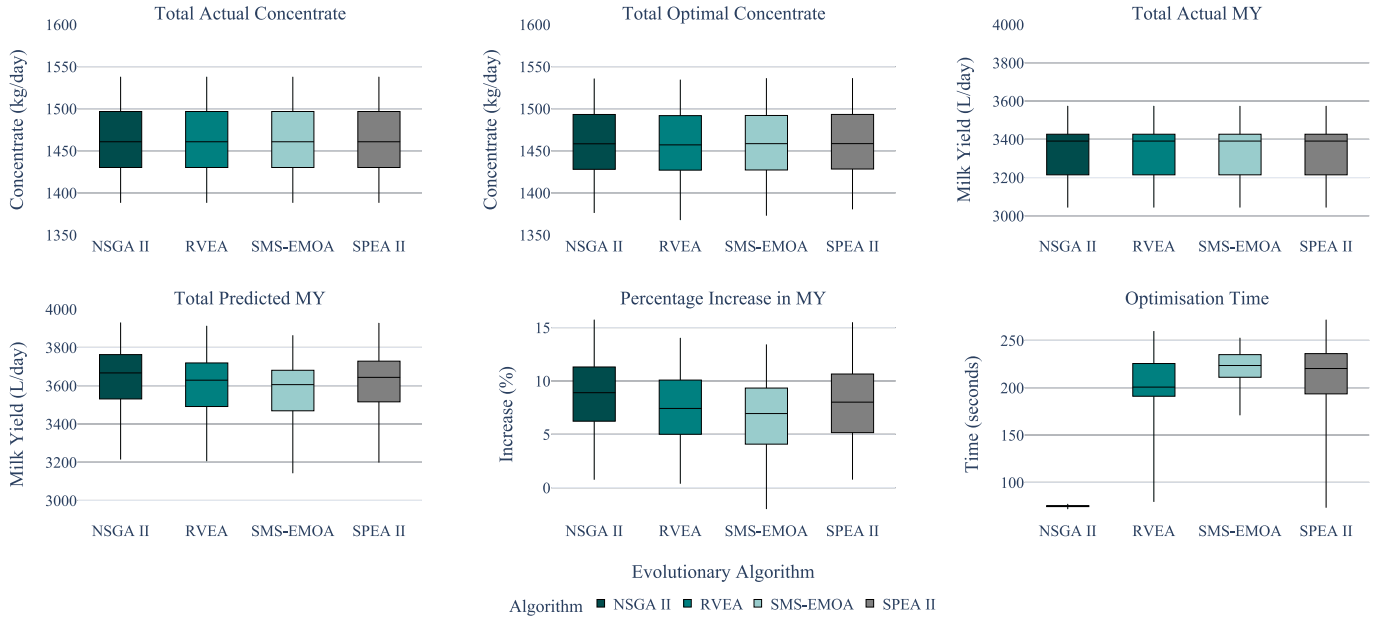


Fig. 8. Daily optimisation performance metrics comparison across evolutionary algorithms.
The boxplots display the distribution of daily performance metrics across 10 independent runs (different random seeds) for each algorithm, with each boxplot representing approximately 300 daily observations (10 runs × 30 days per algorithm). The six panels show: (a) total actual concentrate allocated per day (kg/day), (b) total optimal concentrate per day (kg/day), (c) total actual milk yield per day (L/day), (d) total predicted milk yield per day (L/day), (e) percentage increase in milk yield per day (%), and (f) optimisation time per day (seconds). The boxplots show the median (central line), interquartile range (box), and whiskers extending to 1.5× the interquartile range. Results demonstrate consistent daily performance within algorithms and significant differences between algorithms across the 30-day optimisation period, as confirmed by statistical analysis (see Supplementary Table S3).

can reliably predict MY from cow characteristics, even without traditionally essential data such as liveweight, while the embedded evolutionary optimisation techniques effectively adjust concentrate levels. The model performed comparably to our earlier flat-rate feeding study, and its deployment on commercial farms with computerised bail feeders proved feasible despite pronounced day-to-day variability in concentrate offered per cow. Although computational efficiency remains an area for improvement, the inherent variability in concentrate consumption has, in fact, provided the model with a richer dataset from which to learn, enhancing its ability to prescribe optimal feeding levels. In the flat-rate study the narrow spread in concentrate intake restricted the scope for exploring alternative feeding–response combinations, and the optimisation routine implemented with SciPy was computationally expensive and slow to reach a solution.

In this present work, linear programming approaches has been replaced with evolutionary algorithms, which explores the search space more thoroughly, converges faster, and ultimately yields more effective feeding recommendations. The addition of environmental variables such as daily maximum and minimum temperature, together with a broader range of cow characteristics including breed, has further improved model robustness and its capacity to capture the factors that drive feed efficiency and MY. This outcome suggests that while incorporating additional variables like liveweight and health related information could further refine the process, the current approach is robust enough to deliver reliable predictions with a strategically selected set of input variables.

A key limitation of this study is the absence of individual forage intake and pasture nutritional value data in the predictive model. In pasture-based systems, individual cow forage intake cannot be measured without specialised equipment, limiting our ability to account for the full nutritional profile affecting MY. While our model achieved reliable predictions using concentrate intake and other readily available variables, future research should continue to refine these techniques by incorporating forage quality metrics (e.g., periodic pasture sampling), a broader range of cow characteristics (such as liveweight and health-related information), and environmental data. Additionally, validating the framework across multiple farms with varying forage types, nutritional compositions, and management systems would strengthen generalisability and further enhance the scalability and applicability of such optimisation models in real-world dairy systems.

Whilst the GA identified an optimal 30-day window based on cow representation and DIM balance, we acknowledge that results may exhibit some sensitivity to the specific window selected. Factors such as seasonal pasture quality, environmental conditions, and herd DIM distribution could influence outcomes across different periods. Algorithm rankings are likely to remain consistent across windows, as they reflect inherent algorithmic characteristics rather than seasonal factors. However, the absolute magnitude of milk yield improvements could vary with baseline herd productivity and pasture availability. Future research should validate this approach across multiple seasonal windows and farm systems to confirm the consistency of algorithm performance and characterise the range of potential improvements.

4. Conclusion

This study presents a significant advancement in optimising grain-based concentrate allocation for maximising MY in pasture-based dairy cows. It integrates multi-objective evolutionary algorithms, tailored to computerised individual-feeding systems, with ML models. Unlike earlier work that relied on flat-rate regimes and single, computationally heavy optimisation models our approach balances computational efficiency with performance improvements in commercial farms with individualised feeding capacities. It also differentiates itself by systematically comparing several evolutionary frameworks on data collected in a live farm environment, something absent from earlier flat-rate feed optimisation studies. Our findings indicate that, while there are inevitable trade-offs between optimisation time and MY improvements, coupling the machine learning trained model with evolutionary algorithms effectively tailored concentrate levels to individual cow. Predictions from the MY model act as the objective function for the optimiser, ensuring that concentrate is allocated, thereby maximising daily MY and contributing to farm profitability.

The results show that GBM was the best predictive model, explaining up to 61% of the variability in MY with an RMSE of 3.47 L/cow.day^{-1}. This model developed using cow-level data together with environmental variables, effectively predicts MY, highlighting the importance of machine learning techniques in dairy farming. The study evaluated four multi-objective evolutionary algorithms (NSGA II, RVEA, SMS-EMOA, SPEA II), which boosted MY between 6.63 ± 0.24% and 8.64 ± 0.10%, with all algorithms showing statistically significant differences (ANOVA, $F = 303.24$, $p < 0.001$). Among these, NSGA-II demonstrated the best overall performance, achieving both the highest milk yield increase (8.64 ± 0.10%) and the fastest computational speed (74.9 ± 0.7 s/day), while SPEA-II offered a balance between performance (7.94 ± 0.13% increase) and moderate computational requirements, providing practical advantages for on-farm decision-making.

Recent studies increasingly emphasise that supplement allocation work best when they are customised to each cow, even within herd-level management systems. Our findings reinforce this view: accounting for cow-specific traits and on-farm environmental conditions has shown that it improves the precision of feed recommendations. By integrating those individual-level data into the optimisation framework, we demonstrate tangible gains in both the scalability and accuracy of model outputs on commercial dairy farms. In addition, the computational refinements applied to the SPEA-II algorithm illustrate how targeted algorithmic adjustments and upgrades can further streamline future feed-optimisation research in dairy systems.

In summary, the approach outlined in our study offers a promising direction for optimising dairy feed, maximising milk yield and boosting profitability by leveraging individual cow variability and environmental factors. However, further work is required to validate these models under real farm environments, incorporating additional variables such as pasture utilisation and monitoring cow behaviour in response to concentrate prescriptions. These insights would refine the optimisation process and advance the development of more sustainable, efficient dairy systems.

Author ORCIDs

- **Blessing Azubuike:** https://orcid.org/0000-0001-7902-6155.
- **Anna Chlingaryan:** https://orcid.org/0000-0001-9332-7886.
- **Martin Correa-Luna:** https://orcid.org/0000-0002-1122-0258.
- **Cameron Clark:** https://orcid.org/0000-0002-7644-2046.
- **Sergio C. Garcia**: https://orcid.org/0000-0003-1630-3543.

CRediT authorship contribution statement

Blessing Nnenna Azubuike: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Chlingaryan:** Writing – review & editing, Methodology, Conceptualization. **Martin Correa-Luna:** Supervision, Methodology, Conceptualization. **Cameron E.F. Clark:** Writing – review & editing, Supervision, Conceptualization. **Sergio C. Garcia:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Funding statement

This research was funded by the Dairy UP program, a collaborative RD&E program for New South Wales, Australia (www.dairyup.com.au) through the academic scholarship awarded to Blessing N. Azubuike.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the research scholarship granted by the DairyUP project at the University of Sydney.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiia.2026.04.003.

References

Astuti, V., Raj, K.H., 2016. Hybrid evolutionary computational algorithm for dairy cattle feed cost optimization. 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). https://ieeexplore.ieee.org/document/7906780.

Azubuike, B., Chlingaryan, A., Correa-Luna, M., Clark, C., Garcia, S., 2024. A Data-Driven Approach for Optimising Supplement Allocation to Individual Lactating Dairy Cows in Pasture-Based Systems. Australasian Dairy Science Symposium, ChristChurch, New Zealand (2024, 25 -28 November).

Berry, D.P., Buckley, F., Dillon, P., 2007. Body condition score and live-weight effects on milk production in irish Holstein-Friesian dairy cows. Anim. Int. J. Anim. Biosci. 1 (9), 1351–1359. https://doi.org/10.1017/S1751731107000419.

Blank, J., Deb, K., 2020. Pymoo: multi-objective optimization in python. IEEE Access 8, 89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567.

Bock, H.-H., 2007. Clustering methods: A history of k-means algorithms. In: Brito, P., Cucumel, G., Bertrand, P., de Carvalho, F. (Eds.), Selected Contributions in Data Analysis and Classification. Springer Berlin Heidelberg, pp. 161–172. https://doi.org/10.1007/978-3-540-73560-1_15.

Brady, E.L., Kelly, E.T., Lynch, M.B., Fahey, A.G., Pierce, K.M., Mulligan, F.J., 2022. The effect of concentrate feeding strategy and dairy cow genotype on milk production, pasture intake, body condition score and metabolic status under restricted grazing conditions. Livest. Sci. 256, 104815. https://doi.org/10.1016/j.livsci.2021.104815.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Campos, L.M., Ringer, H., Chung, M., Hanigan, M.D., 2023. Application of a mathematical framework for the optimization of precision-fed dairy cattle diets. Animal 17, 101001. https://doi.org/10.1016/j.animal.2023.101001.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 https://doi.org/10.1145/2939672.2939785.

Cockburn, M., 2020. Review: application and prospective discussion of machine learning for the management of dairy farms. Animals 10 (9), 1690. https://www.mdpi.com/2076-2615/10/9/1690.

Craig, A.-L., Gordon, A.W., Hamill, G., Ferris, C.P., 2022. Milk composition and production efficiency within feed-to-yield systems on commercial dairy farms in Northern Ireland. Animals 12 (14), 1771. https://doi.org/10.3390/ani12141771.

Dale, A.J., McGettrick, S., Gordon, A.W., Ferris, C.P., 2016. The effect of two contrasting concentrate allocation strategies on the performance of grazing dairy cows. Grass Forage Sci. 71 (3), 379–388. https://doi.org/10.1111/gfs.12185.

Das, R., Nath Das, K., Mallik, S., 2023. An efficient evolutionary optimizer for solving complex dairy feed optimization problems. Comput. Electron. Agric. 204, 107566. https://doi.org/10.1016/j.compag.2022.107566.

Deb, K., 2001. Multiobjective Optimization Using Evolutionary Algorithms. Wiley, New York.

Dela Rue, B.T., Eastwood, C.R., 2017. Individualised feeding of concentrate supplement in pasture-based dairy systems: practices and perceptions of New Zealand dairy farmers and their advisors. Anim. Prod. Sci. 57 (7), 1543–1549. https://doi.org/10.1071/AN16471.

Feeding dairy cows: a manual for use in the Target 10 Nutrition Program. (2015). In J. Jacobs & A. Hargreaves (Eds.), (5th ed. ed., pp. 14.11-14.16). Warrnambool, Vic.:: Dept. of Natural Resources and Environment.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gasser, L., Cruz, F.P., Cockburn, M., 2023. Can meteorological data improve the short-term prediction of individual milk yield in dairy cows? J. Dairy Sci. 106 (8), 5501–5516. https://doi.org/10.3168/jds.2022-22980.

Henriksen, J.C.S., Weisbjerg, M.R., Løvendahl, P., Kristensen, T., Munksgaard, L., 2019. Effects of an individual cow concentrate strategy on production and behavior. J. Dairy Sci. 102 (3), 2155–2172. https://doi.org/10.3168/jds.2018-15477.

Hills, J.L., García, S.C., Dela Rue, B., Clark, C.E.F., 2015. Limitations and potential for individualised feeding of concentrate supplements to grazing dairy cows. Anim. Prod. Sci. 55 (7), 922–930. https://doi.org/10.1071/AN14855.

Ji, B., Banhazi, T., Phillips, C.J.C., Wang, C., Li, B., 2022. A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. Biosyst. Eng. 216 (9), 186–197. https://doi.org/10.1016/j.biosystemseng.2022.02.013.

Katsini, L., Muñoz López, C.A., Bhonsale, S., Roufou, S., Griffin, S., Valdramidis, V., Akkermans, S., Polanska, M., Van Impe, J., 2024. Modeling climatic effects on milk production. Comput. Electron. Agric. 225, 109218. https://doi.org/10.1016/j.compag.2024.109218.

Lee, M., Lee, S., Park, J., Seo, S., 2020. Clustering and characterization of the lactation curves of dairy cows using K-Medoids clustering algorithm. Animals 10 (8), 1348. https://doi.org/10.3390/ani10081348.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2 (3), 18–22.. https://journal.r-project.org/articles/RN-2002-022/.

Mason, L., Baxter, J., Bartlett, P., Frean, M., 1999. *Boosting Algorithms as Gradient Descent* Proceedings of the 12th International Conference on Neural Information Processing Systems, Denver, CO. https://doi.org/10.5555/3009657.3009730.

Nguyen, Q.T., Fouchereau, R., Frénod, E., Gerard, C., Sincholle, V., 2020. Comparison of forecast models of production of dairy cows combining animal and diet parameters. Comput. Electron. Agric. 170, 105258. https://doi.org/10.1016/j.compag.2020.105258.

Notte, G., Cancela, H., Pedemonte, M., Chilibroste, P., Rossing, W.A.H., Groot, J.C.J., 2020. A multi-objective optimization model for dairy feeding management. Agric. Syst. 183, 102854. https://doi.org/10.1016/j.agsy.2020.102854.

Notte, G., Chilibroste, P., Pedemonte, M., Cancela, H., 2021. *Evolutionary multi-objective algorithms for feed resource allocation in dairy systems* [Conference paper]. 2021 IEEE latin american conference on computational intelligence, LA-CCI 2021. https://ieeexplore.ieee.org/document/9769787.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G., 2012. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12. https://doi.org/10.48550/arXiv.1201.0490.

Shu, H., Li, Y., Bindelle, J., Jin, Z., Fang, T., Xing, M., Guo, L., Wang, W., 2023. Predicting physiological responses of dairy cows using comprehensive variables. Comput. Electron. Agric. 207, 107752. https://doi.org/10.1016/j.compag.2023.107752.

Souza, V.C., Liebe, D.M., Price, T.P., Ellett, M.D., Davis, T.C., Gleason, C.B., Daniels, K.M., White, R.R., 2022. Algorithm development for individualized precision feeding of supplemental top dresses to influence feed efficiency of dairy cattle. J. Dairy Sci. 105 (5), 4048–4063. https://doi.org/10.3168/jds.2021-20841.

Streefland, G.-J., Herrema, F., Martini, M., 2023. A gradient boosting model to predict the milk production. Smart Agricul. Technol. 6, 100302. https://doi.org/10.1016/j.atech.2023.100302.

Usigbe, M.J., Darlan, D., Uyeh, D.D., Mallipeddi, R., 2023. *Animal Feed Optimization under Price Fluctuations Using Evolutionary Algorithms* 2023 14th International Conference on Information and Communication Technology Convergence (ICTC). https://ieeexplore.ieee.org/document/10393678.

Uyeh, D.D., Mallipeddi, R., Pamulapati, T., Park, T., Kim, J., Woo, S., Ha, Y., 2018. Interactive livestock feed ration optimization using evolutionary algorithms. Comput. Electron. Agric. 155, 1–11. https://doi.org/10.1016/j.compag.2018.08.031.

Uyeh, D.D., Pamulapati, T., Mallipeddi, R., Park, T., Asem-Hiablie, S., Woo, S., Kim, J., Kim, Y., Ha, Y., 2019. Precision animal feed formulation: an evolutionary multi-objective approach. Anim. Feed Sci. Technol. 256, 114211. https://doi.org/10.1016/j.anifeedsci.2019.114211.

Zhang, F., Upton, J., Shalloo, L., Shine, P., Murphy, M.D., 2020. Effect of introducing weather parameters on the accuracy of milk production forecast models. Inform. Process. Agricul. 7 (1), 120–138. https://doi.org/10.1016/j.inpa.2019.04.004.

CHAPTER 5

Data augmentation and interpolation improves machine learning-based pasture biomass estimation from sentinel-2 imagery

Chapters 3 and 4 showed that milk yield can be improved by optimising how concentrate is allocated to individual cows, but concentrate is only part of the diet. The cheapest feed source on any pasture-based dairy farm is the pasture itself, and it remains the most difficult to monitor. Pasture is typically three to five times cheaper than grain-based concentrate when managed correctly, yet profitable pasture management depends entirely on knowing how much is available at any given time. Satellite remote sensing offers a promising pathway for scalable, continuous pasture biomass monitoring across large farm areas, yet frequent cloud cover renders it unreliable for continuous assessment, and manual rising plate meter measurements are impractical at the scale required for daily farm management decisions. To overcome these limitations, we integrate Sentinel-2 imagery, ground truth measurements and weather data with interpolation methods to generate essential temporal data continuity. This approach significantly enhanced predictive accuracy for pasture biomass estimation, filling critical data gaps and establishing a vital tool that unlocks the economic potential of this primary feed resource. All data preprocessing and modelling implementations are available in two Jupyter notebooks in the project repository (data_cleaning.ipynb and modelling_visualisations.ipynb) with key code excerpts are presented in Appendix C.

Remote Sensing (2025) 17(23), 3787

PUBLISHED MANUSCRIPT

The published version of this manuscript is included on the following page.

remote sensing

MDPI

Article

Data Augmentation and Interpolation Improves Machine Learning-Based Pasture Biomass Estimation from Sentinel-2 Imagery

Blessing N. Azubuike [1,2,*], Anna Chlingaryan [2,3], Martin Correa-Luna [1,2], Cameron E. F. Clark [2,4] and Sergio C. Garcia [1,2]

[1] Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW 2570, Australia
[2] Dairy UP Program, Camden, NSW 2570, Australia; sergio.garcia@sydney.edu.au (S.C.G.)
[3] Livestock Production and Welfare Group, School of Life and Environmental Sciences, The University of Sydney, Camden, NSW 2570, Australia
[4] Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW 2650, Australia; camclark@csu.edu.au
* Correspondence: blessing.azubuike@sydney.edu.au

Highlights

What are the main findings?

- Full-band Sentinel-2 reflectance combined with weather variables substantially improved pasture biomass prediction accuracy, outperforming vegetation indices and achieving up to 70% variance explained after interpolation.
- Multiquadric interpolation and progressive temporal training strengthened temporal consistency in the data, reducing prediction errors by approximately 30% relative to sparsely sampled baselines.

What are the implications of the main findings?

- These findings demonstrate that integrating physically meaningful spectral information with biologically constrained data augmentation enhances the reliability and scalability of satellite-based biomass estimation across farms and seasons.
- The resulting open-source modelling framework provides a robust foundation for real-time pasture monitoring and can be readily incorporated into automated decision-support tools for precision grazing management.

check for updates

Academic Editors: Christoph Jörges and Aaron Moody

Received: 13 October 2025
Revised: 11 November 2025
Accepted: 13 November 2025
Published: 21 November 2025

Citation: Azubuike, B.N.; Chlingaryan, A.; Correa-Luna, M.; Clark, C.E.F.; Garcia, S.C. Data Augmentation and Interpolation Improves Machine Learning-Based Pasture Biomass Estimation from Sentinel-2 Imagery. *Remote Sens.* **2025**, *17*, 3787. https://doi.org/10.3390/rs17233787

Abstract

Accurate pasture biomass (PB) estimation is critical for tactical grazing management, yet traditional satellite-derived vegetation indices such as Normalised Difference Vegetation Index (NDVI) saturate when canopy density exceeds about 3 t DM ha^{-1}. This limits predictive accuracy because the spectral signal plateaus under dense vegetation, masking further biomass increases. To address this limitation, this study integrated multiple data sources to improve PB estimation in dairy systems. The dataset combined Sentinel-2 spectral bands, rising plate-meter (RPM) PB measurements, daily weather data, and paddock management features. A total of 3161 paired RPM–satellite observations were collected from 80 paddocks across 16 New South Wales dairy farms between November 2021 and July 2024. Eight regression algorithms and four predictor configurations were evaluated using robust cross-validation, including an 80:20 farm/paddock-stratified train–test-set split. The XGBoost model using full-band reflectance and concurrent weather data achieved strong baseline performance (R^2 = 0.63; MAE = 243 kg DM ha^{-1}) on non-interpolated data, outperforming NDVI-based models. To address temporal gaps between field readings and

satellite imagery, Multiquadric interpolation was applied to RPM data, adding roughly 30% new observations. This enhanced dataset improved test performance to R^2 = 0.70 and MAE = 216 kg DM ha^{-1}, with gains maintained on external validations (R^2 = 0.41/0.48; MAE = 267/235 kg DM ha^{-1}). A progressive training strategy, which refreshed model parameters with seasonally aligned data, further reduced errors by 30% compared to static models and sustained performance even when farms or seasons were excluded. This fortified Sentinel-2 modelling workflow, combining RPM interpolation and progressive calibration, achieved accuracy comparable to the commercial Pasture.io platform (R^2 = 0.66; MAE = 240 kg DM ha^{-1}) which uses satellite imagery with higher temporal and spatial resolution, demonstrating potential for automated recalibration and near real-time, paddock-level decision support in pasture-based dairy systems.

Keywords: dairy systems; data augmentation; data interpolation; machine learning; pasture biomass estimation; remote sensing; Sentinel-2

1. Introduction

Reliable estimation of above-ground pasture biomass or pasture biomass (PB) is essential for effective farm management, feed budgeting, and decision-making in pasture-based dairy systems. As global demand for agricultural efficiency grows, accurate and scalable PB monitoring tools become increasingly important. Pasture biomass estimation supports livestock system productivity, maintains environmental sustainability through optimised resource use, and improves economic viability across diverse farming operations [1]. However, conventional data collection methods like rising plate meters (RPM), a tool that relates compressed height of PB with the amount of biomass present, provide relatively high-quality data but are labour intensive, time consuming and impractical for daily monitoring, and this could create operational bottlenecks for farmers seeking to balance productivity with operational sustainability [2,3].

In contrast, satellite remote sensing provides broad scale, repeatable and comparatively low cost coverage of vegetation condition [1,4]. Yet, its most widely used proxy, the Normalised Difference Vegetation Index (NDVI), saturates under dense canopies, reducing sensitivity to PB and flattening response curves, which compromises predictive accuracy [5,6]. Alternative indices, including the Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI) and Normalised Difference Red-Edge Index (NDRE), are derived from different combinations of spectral bands and offer only marginal improvements, unable to accurately predict PB on their own [7,8]. Because PB variability is driven by multiple factors such as soil moisture, paddock management, short-term weather extremes and region-specific sward composition, indices alone cannot explain PB variation; these additional drivers must be represented explicitly in modelling frameworks [9,10].

Sentinel-2 provides free, high-temporal and high-spatial resolution multispectral imagery that supports large-scale, frequent PB measurements, making it a cost-effective alternative relative to other options like UAV-mounted sensors, field spectrometers, and ground-based cameras [10–13]. Its diverse spectral bands allow researchers to monitor vegetation health and growth consistently and affordably [14,15].

However, satellite-based data utilisation has its inherent limitations [16]. Cloud cover, atmospheric interference and inclement weather conditions can produce unreliable satellite-derived products or no data at all, limiting PB estimates for decision-making using this approach and reducing reliable insights on pasture management and utilisation [4,8]. On days when satellite data are completely unavailable, for example on cloudy or snowy

days, ground-based RPM measurements are usually employed to compensate for missing satellite data to maintain consistent flow of data [9,10,17,18].

Integrating diverse data sources offers a potential solution to these challenges. Recent research shows that combining raw satellite imagery with ground-truth measurements, weather variables and paddock-specific information, can significantly enhance PB prediction accuracy [10,18]. Machine learning (ML) algorithms such as Random Forests (RF), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been used in these integrations and outperform traditional regression methods and linear programming by providing better insights, identifying nonlinear relationships between driving factors affecting pasture growth and managing high-dimensional data, and adapting to temporal and spatial variability, thereby offering a more comprehensive understanding of pasture utilisation and management [19–22].

As an alternative or complement to ML approaches, physiological growth models provide mechanistic insights into pasture dynamics by simulating processes such as photosynthesis, respiration, and nutrient uptake based on environmental drivers. Models such as DairyMod, APSIM-Pasture, and ModVege have been successfully applied to predict pasture growth in temperate grazing systems, offering interpretable predictions grounded in plant ecophysiology [23–26]. While these process-based models excel at capturing temporal growth patterns under varying environmental conditions, they typically require extensive parameterisation and may lack the flexibility to incorporate high-dimensional remote sensing data directly [27]. Hybrid approaches that combine the mechanistic foundation of physiological models with the pattern-recognition capabilities of ML algorithms represent a promising research direction, potentially leveraging the strengths of both methodologies to improve predictive accuracy and model interpretability [28,29].

However, significant research gaps persist, particularly regarding temporal and spatial misalignments between satellite data and ground measurements, which create inconsistencies as satellite observations often do not coincide with the timing or location of ground data collection. Furthermore, missing data due to cloud cover or other environmental obstructions disrupts the continuity of satellite-derived datasets [4,30]. Addressing these issues requires sophisticated data manipulation strategies such as interpolation techniques to fill gaps, synthetic data generation to expand datasets and progressive training to capture temporal patterns across different scenarios [31]. The generalisability of models across regions and farming systems is also a concern as a model trained on data from one area or region may perform well within that context, however, its applicability to other areas or locations with different weather, soil, and management conditions is not guaranteed. Therefore, evaluating reliability and flexibility using external validation datasets is essential to ensure that the predictive framework is not overly tailored to a specific environment but remains versatile enough for broader adoption.

Building on these challenges highlighted, this research poses several key questions to address the limitations of current pasture estimation methods. First, how can the integration of raw satellite imagery, RPM measurements, weather data, and paddock-specific characteristics significantly enhance PB prediction accuracy, especially overcoming the known saturation issues of conventional derived vegetation indices like NDVI? Second, how effectively can the integration of diverse data sources address the challenges of missing or unreliable satellite readings caused by cloud cover, thereby ensuring data continuity for robust model performance? Third, can advanced ML models effectively compensate for temporal and spatial mismatches, ensuring their validity and reliability when applied to external datasets across varied weather, soil, and management contexts, thereby offering a robust and generalisable solution beyond existing site-specific approaches?

To address these questions, the objectives of this study are threefold. First, to develop an integrated ML framework that leverages raw Sentinel-2 reflectance, RPM-derived PB, weather, and paddock management characteristics to accurately estimate PB. Second, to evaluate the reliability and transferability of the framework across diverse temporal and spatial conditions, including unseen farms and seasons, by leveraging data augmentation and progressive training strategies. Third, to benchmark the developed framework against a commercial platform, demonstrating its practical application as a transparent, cost-effective and scalable solution for sustainable near real-time, farm-specific pasture utilisation and management in dairy systems.

2. Materials and Methods

2.1. Study Area and Data Sources

This study was conducted across 16 commercial dairy farms located in three coastal districts of New South Wales (NSW), Australia, between November 2021 and July 2024 (Figure 1). The farms were distributed across the mid-coast district (n = 7 farms, latitude range: −34.03° to −31.72°S, longitude range: 150.65° to 152.68°E), the south coast (n = 5 farms, latitude range: −36.82° to −36.64°S, longitude range: 149.60° to 149.90°E), and the north coast (n = 4 farms, latitude range: −28.90° to −28.68°S, longitude range: 152.91° to 153.13°E), with specific farm coordinates withheld to maintain commercial confidentiality. These farms collectively managed 2436 hectares of grazing land, with herd sizes varying from 105 to 580 milking cows, and individual farms containing between 22 and 83 paddocks, with utilisable grazing areas ranging from 64.5 ha to 313 ha. The study regions presented diverse weather conditions, with long-term mean annual rainfall averaging approximately 780 mm in the south coast, 1284 mm in the mid-coast, and 1073 mm in the north coast. Average daily air temperatures ranged from approximately 5 to 20 °C in winter and 20 to 35 °C in summer. All farms had pastures based on kikuyu (*Cenchrus clandestinus*, previous *Pennisetum clandestinum*), which produces biomass from late spring (November) through summer and autumn, oversown every year (March to April) with annual ryegrass (*Lolium multiflorum* L.), which produces biomass during autumn, winter, and early spring.

Remote sensing data was acquired from Copernicus Sentinel-2 surface reflectance imagery, which offers 10 m spatial resolution and a nominal five-day revisit cycle for each paddock in each farm during clear-sky passes. The spectral bands retained for analysis included blue, green, red, near-infrared, red-edge 1 to 3, and short-wave infrared 2 and 3. Standard vegetation indices such as NDVI, EVI, SAVI, and NDRE were also calculated per pixel. To ensure data quality, pixels were masked using the Function of mask (Fmask) layer, classifying them as valid, water, cloud, shadow, or snow.

Ground-truth PB measurements were conducted using an RPM. A Jenquip EC20 Electronic Pasture Meter (Feilding, New Zealand) was specifically utilised to obtain Compressed Sward Height (CSH) readings. Five primary paddocks on each farm were selected for continuous monitoring over the two-year study period (November 2021 to July 2024), along with one additional reserve paddock (the sixth paddock) which served as a spatially independent validation site. The selection of these five primary paddocks was based on the following criteria: (i) representativeness of the predominant pasture species composition on each farm (kikuyu oversown with annual ryegrass), (ii) accessibility for consistent fortnightly/weekly sampling throughout the study period, (iii) diversity in paddock size and topographic characteristics to capture farm-level variability, and (iv) active incorporation in the farm's regular grazing rotation to ensure realistic management conditions. The sixth paddock was intentionally withheld from model training to provide an independent spatial validation dataset, representing genuine out-of-sample conditions for assessing model generalisation to unseen paddock locations within the same farm.

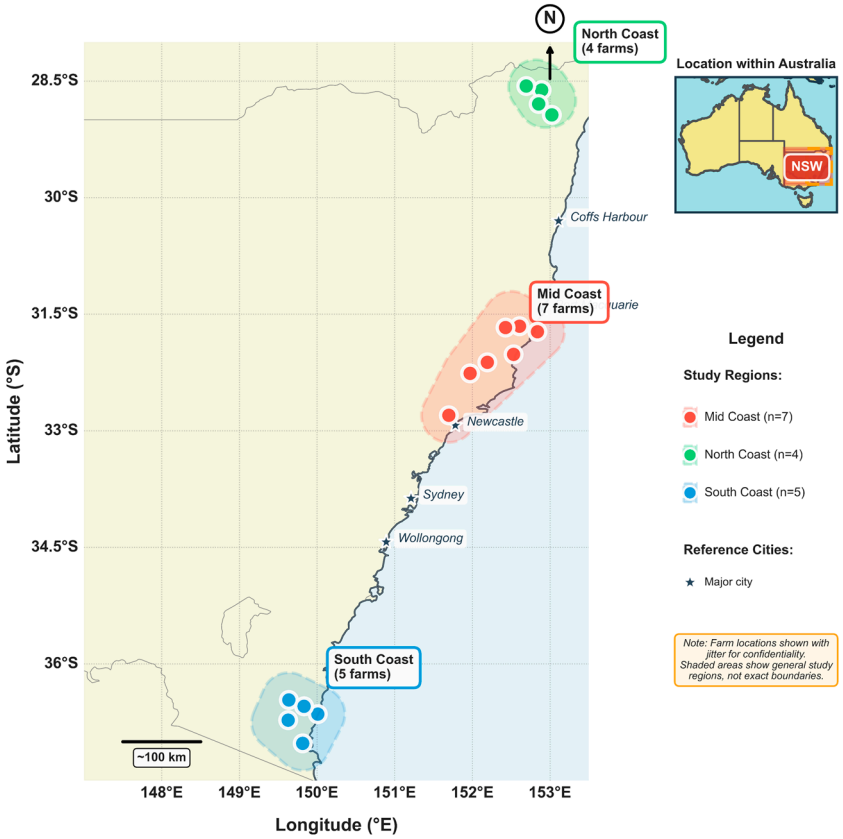


Figure 1. Geographic distribution of study farms across coastal districts of New South Wales, Australia used in the study. This illustrates the spatial distribution of the 16 commercial dairy farms included in this study, located across three coastal districts of New South Wales (NSW), Australia. Farms were grouped into the mid-coast (n = 7), south coast (n = 5), and north coast (n = 4) regions, as defined in Section 2.1. Marker positions represent jittered farm centroids derived from paddock-level coordinates, with a small random displacement applied to protect commercial confidentiality. Regional shading represents the general geographic coverage of each study district. An inset map of Australia highlights the location of NSW relative to the broader states.

RPM calibration was an integral part of the data collection, performed monthly on two designated fixed paddocks per farm [10]. The calibration process involved collecting nine 0.1 m^2 quadrat cuts at a 5 cm stubble height, stratified across the paddock to capture high, medium, and low biomass zones (three cuts per zone). Samples were dried for 48 hours at 65 °C, weighed, and subsequently regressed against the corresponding CSH measurements to derive farm-specific and seasonally adjusted conversion equations. This approach ensured that CSH-to-PB conversions accounted for seasonal changes in pasture density and species composition. The monthly calibration frequency was designed to capture phenological changes in sward structure that could affect the CSH-PB relationship, particularly during transitions between kikuyu and ryegrass dominance. This process allowed for the accurate conversion of CSH readings into PB, expressed as kilograms of Dry Matter per hectare (kg DM ha^{-1}). For each paddock, a minimum of 70 plate measurements were recorded and then converted to PB in kg DM ha^{-1} [10,18]. The spatial distribution of measurements within each paddock ensured that areas with varying slope, drainage, and proximity to high-traffic zones (gates, water points, shade structures) were adequately represented, thereby minimising bias from localised management effects or microtopographic variation.

Additionally, environmental data consisting of daily weather variables for each farm were obtained from the SILO Long Paddock platform (https://www.longpaddock.qld.gov.au).

Variables used were maximum air temperature (°C), minimum air temperature (°C), rainfall (mm), vapour pressure (kPa), maximum and minimum relative humidity (%), incoming solar radiation (MJ m^{-2}) and evapotranspiration (mm). These observations were obtained to provide essential information regarding weather influences on PB.

2.2. Data Preprocessing

Sentinel-2 images were corrected to surface reflectance, resampled to a common 10 m grid and clipped to paddock boundaries obtained from the georeferenced GIS dataset for each farm. The pixel-quality layer Fmask [32] classified each pixel as valid, water, cloud, shadow or snow, and only pixels flagged valid were retained for analysis. For every monitored paddock on the day an image was taken, the mean reflectance of blue, green, red, near-infrared, red-edge 1-3 and short-wave-infrared 2-3 was calculated, and vegetation indices NDVI, EVI, SAVI and NDRE were derived. Spectral outliers falling outside 1.5 interquartile ranges from the first or third quartile were removed before averaging.

Rising plate-meter records were filtered to keep PB between 1000 kg DM ha^{-1} and 4000 kg DM ha^{-1}. Approximately 70 individual RPM readings collected within each paddock on a monitoring day were averaged to one PB value per paddock. Within each farm-paddock-date group, categorical fields such as paddock name were reduced to their modal value and numeric fields were averaged. Daily SILO weather data were converted to numeric, checked for implausible entries and merged directly by each farm and calendar date.

Date stamps in every dataset were converted to datetime and expanded to ISO week number, calendar year and austral seasons, where Spring includes September to November, Summer (December to February), Autumn (March to May) and Winter (June to August). Satellite data, daily weather observations and PB estimates were merged on farm code, paddock code and date. When Sentinel and plate-meter observations were not recorded on the same day, PB rows were retained only if a Sentinel acquisition occurred within $\pm$3days. Rows with missing predictor values were removed on a complete-case basis, and continuous variables were centred and scaled to unit variance for subsequent modelling, yielding a curated dataset of 3161 records from 80 paddocks across the 16 farms.

2.3. Interpolation Methods for Data Augmentation

Following the merging of the RPM measurements, collected weekly in year 1 and fortnightly in year 2, with Sentinel-2 passes that recur on an approximately five-day revisit cycle, small temporal gaps remained in PB because satellite imagery does not provide a direct PB measurement from the bands. To expand the dataset, these gaps were filled within each paddock in each farm using stochastic interpolation routines implemented in Python (v3.11.4).

Four interpolation techniques were assessed to augment PB observations for the merged dataset (RPM and Sentinel). The set comprised a second order polynomial in time as the baseline curve, a Gaussian radial basis function, a multiquadric radial basis function [33,34] and a minimum curvature exact spline. Radial basis surfaces offered flexible data-driven alternatives that perform well for smooth, yet non-linear environmental series [35–37] and the minimum curvature spline minimises the surface Laplacian, a property valued in geophysical analysis [38]. All four algorithms were executed independently three times for each paddock in each farm: once on year 1 data only (dates earlier than 1 April 2023), once on year 2 data only (dates on or after 1 April 2023) and once on the combined two-year record to exploit longer term temporal structure where available.

Interpolation was performed only when at least three actual PB observations were available for a paddock and was strictly confined to the temporal span defined by those

observations, thereby preventing any extrapolation. Analysis of observation intervals revealed that time gaps between consecutive pasture meter measurements in the final dataset ranged from 7 to 14 days (median = 7 days, mean = 8.4 days). The majority of intervals (84.9%) were $\leq$14 days, with 94.3% $\leq$30 days and only 1.6% exceeding 60 days. The multiquadric radial basis function employed is appropriate for capturing smooth temporal trajectories across such intervals, as pasture growth follows gradual, predictable patterns between discrete grazing events rather than exhibiting abrupt discontinuities. For the small proportion of longer intervals (>60 days), interpolation remained constrained to the temporal span of observed measurements, with no extrapolation beyond the empirical record. All time stamps were converted to Unix epoch seconds to provide a uniform, monotonic reference axis for curve fitting. Prior to modelling, any PB measurement falling outside the biologically credible interval of 1000 to 4000 kg DM ha^{-1} was discarded. After the gap-filling procedure each record was annotated to indicate whether Sentinel-2 reflectance originated from a nearest-date substitution and whether the corresponding PB simulated was replaced by an interpolated value, enabling downstream analyses to distinguish directly measured data from synthetically generated values. Rows with null values after processing were discarded. The combination of biological filtering, nearest-date filling and advanced interpolation produced a comprehensive temporal gap-free dataset containing 9816 daily records from 80 paddocks across the 16 farms.

2.4. Predictive Modelling for Pasture Biomass

2.4.1. Model Training and Optimisation

Univariate analysis was employed to quantify linear relationships between each predictor and the response variable PB (kg DM ha^{-1}) using a Pearson correlation matrix visualised as a heat map. Numerical predictors comprised daily maximum temperature (°C), daily minimum temperature (°C), evapotranspiration (mm d^{-1}), incoming solar radiation (MJ m^{-2} d^{-1}), vapour pressure (kPa), rainfall (mm d^{-1}), maximum and minimum relative humidity (%), ten (10) Sentinel-2 spectral bands (blue, green, red, near infra-red, red edge 1 to 3, short-wave infra-red 2 and 3) and four derived vegetation indices (NDVI, SAVI, EVI, NDRE). While standard greenness indices (NDVI, EVI, SAVI, NDRE) were calculated, SWIR-based indices such as the Cellulose Absorption Index (CAI) were not explicitly derived, as the tree-based ML algorithms were expected to capture relevant SWIR band information through non-linear feature interactions. Categorical predictors were season, coastal region and grazing information; these were examined with box plots and frequency tables. Four feature sets were defined: (i) all bands with indices, (ii) all bands with weather data and indices, (iii) bands only and (iv) bands with weather data without indices. These configurations were assessed to determine which combination of variables most effectively models PB and to quantify the contribution of weather and vegetation indices to its variability. Predictors showing negligible correlation or high collinearity were removed.

All analyses in this study were carried out in Python (v3.11.4). Categorical variables were one-hot encoded and numerical features were centred and scaled with StandardScaler. Data were split randomly into training and test partitions at an 80:20 ratio, after which multiple random seeds were evaluated through a five-fold three-repeat Repeated K-Fold cross-validation loop to stabilise estimates. Eight regression algorithms, linear regression (LR), least absolute shrinkage and selection operator (LASSO), decision tree (DT), support vector regression (SVR), k-nearest neighbours (KNN), random forest (RF), gradient boosting machine (GBM) and extreme gradient boosting (XGBoost), were wrapped in scikit-learn pipelines. Hyper-parameters for every algorithm were declared in a single Python dictionary; tree ensembles varied the number of decision trees (n_estimators) from 50 to 450 and

maximum depth from 3 to 10, while SVR varied the cost regularisation parameter C from 0.1 to 10 and kernel type. A grid search procedure was employed to systematically test all hyper-parameter combinations, selecting the configuration that minimised negative mean absolute error (MAE, kg DM ha^{-1}), and the model with the lowest cross-validated error was retained.

2.4.2. Progressive Training for Temporal Consistency

To examine how sequential retraining influences model performance while preserving the chronological order of new observations, a progressive training strategy developed by Correa-Luna, et al. [10] was adopted. Every record carried calendar year, month, ISO week, and week of month (WOM), enabling four nested training subsets that represented 25%, 50%, 75% and 100% of each monthly cycle. The subsets were defined as follows: 1 W used WOM = 1; 2 W used WOM = 1 or 2; 3 W used WOM = 1, 2 or 3; and 4 W used WOM = 1, 2, 3 or 4. After each increment, the best model was refitted on the enlarged subset and evaluated on the unchanged test split, with MAE retained as the optimisation metric. The protocol was run separately on the non-interpolated dataset and on the interpolated dataset created in Section 2.3, ensuring that temporal consistency was assessed under both raw and augmented conditions without introducing look-ahead bias. Additionally, to assess inter-annual generalisation, an additional experiment trained the model onYear1Set (data collected before 1 April 2023) and evaluated it on Year2Set (data collected on or after that date), using three dataset variants: the non-interpolated data, a Year 1 multiquadric-interpolated dataset and a full-period multiquadric-interpolated dataset.

2.4.3. Pasture Biomass Model Validation

For the validation, two independent hold-out samples were excluded from all training steps, including non-interpolated, interpolated, and progressive training workflows, and were kept free of any interpolation or gap-filling procedures to represent truly unseen data. The first independent validation sample comprised 41 records representing all available and valid paired RPM PB and Sentinel-2 imagery observations collected from the five primary monitored paddocks across nine of the study farms between 1 and 30 November 2024. The specific count of 41 records was the outcome of applying the data-quality filters (i.e., PB values between 1000 and 4000 kg DM ha^{-1}) to all available measurements during this validation period. For these records, PB measured with RPM was merged with same-day Sentinel-2 imagery, and cloud-affected scenes were handled by replacing them with the corresponding weekly mean reflectance. The selection of these nine farms was based on the availability of complete, high-quality independent data that met the validation criteria during this specific November 2024 period, ensuring they provided fresh, untainted observations.

The second independent validation sample consisted of 63 records reflecting all valid observations gathered specifically from the sixth monitored paddock on each participating farm for the whole period. This "sixth paddock" was intentionally designated as an additional, spatially distinct unseen geographic validation set. The 63 records represent the total number of valid observations obtained after applying the identical merging procedure and the same 1000–4000 kg DM ha^{-1} PB filtering as the training data, ensuring comparable data quality while maintaining their independence. The predictive accuracy of the models on these truly independent datasets was quantified using standard performance metrics: root mean squared error (RMSE), mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). These metrics were calculated separately for each independent hold-out sample, thereby enabling a robust assessment of the ability of the

model to generalise to completely unseen, non-interpolated data across different temporal and spatial contexts.

3. Results

3.1. Dataset Overview and Exploratory Analysis

Descriptive analysis of the final curated non-interpolated paired RPM and Sentinel-2 dataset (3161 observations) resulted in a mean CSH of 80.2 ± 18.1 mm, an average PB of 2690 ± 503 kg DM ha^{-1}, with environmental drivers including daily maximum temperature of 23.7 ± 5.3 °C, rainfall averaging 1.7 ± 7.3 mm d^{-1}, and a mean NDVI of 0.72 ± 0.12 (Table 1). Average PB values exhibited marked seasonal and regional variation across all farms, peaking in summer at approximately 2900 kg DM ha^{-1} and declining to around 2526 kg DM ha^{-1} in winter, representing a 15% amplitude primarily driven by radiation and temperature rather than instantaneous rainfall. Mid-coast paddocks averaged 2830 kg DM ha^{-1}, the south coast 2644 kg DM ha^{-1}, and the north coast 2582 kg DM ha^{-1}.

Table 1. Descriptive statistics for pasture, weather, and spectral variables (n = 3161) used in the study.

Variables	Min.	25%	Median	Mean	75%	Max.	SD
Compressed Height (mm)	24.68	68.02	81.81	80.20	94.45	145.00	18.06
Pasture Biomass (kg DM ha^{-1})	1153.33	2352.08	2737.85	2690.20	3087.74	3720.00	503.06
Daily Max. Temperature (°C)	11.00	19.60	23.30	23.73	27.60	43.00	5.26
Daily Min. Temperature (°C)	−1.10	6.50	11.10	10.97	15.60	27.10	5.66
Rainfall (mm day^{-1})	0.00	0.00	0.00	1.70	0.30	115.20	7.27
Evapotranspiration (mm day^{-1})	0.00	2.10	3.50	3.91	5.40	10.80	2.21
Solar Radiation (MJ m^{-2} day^{-1})	2.90	11.40	14.70	16.46	21.40	31.80	6.80
Vapour Pressure (kPa)	5.40	10.80	14.20	15.20	19.10	29.80	5.33
Max. Relative Humidity (%)	19.40	42.40	49.10	50.33	56.80	93.20	11.87
Min. Relative Humidity (%)	43.70	100.00	100.00	96.77	100.00	100.00	8.22
Blue Band Reflectance	204.91	355.06	421.72	446.73	515.01	836.00	124.39
Green Band Reflectance	379.00	654.69	716.94	735.07	804.70	1117.50	114.56
Red Band Reflectance	214.83	406.97	518.76	564.23	675.00	1307.17	205.89
NIR Band Reflectance	1024.00	3248.95	3758.85	3730.16	4247.21	5996.74	780.05
Red-Edge 1 Reflectance	514.90	986.59	1096.48	1130.08	1257.62	1789.98	214.68
Red-Edge 2 Reflectance	999.93	2554.54	2871.03	2875.78	3231.93	4475.48	544.21
Red-Edge 3 Reflectance	965.17	3010.53	3491.91	3485.73	3987.18	5744.56	739.27
SWIR-2 Reflectance	779.00	1888.85	2112.62	2167.65	2420.06	3478.27	442.39
SWIR-3 Reflectance	327.00	880.23	1028.72	1099.93	1266.77	2200.18	314.58
NDVI	0.27	0.66	0.75	0.72	0.82	0.92	0.12
EVI	0.10	0.47	0.59	0.57	0.69	1.01	0.15
SAVI	0.10	0.43	0.52	0.50	0.59	0.75	0.11
NDRE	0.12	0.45	0.54	0.52	0.60	0.74	0.11

Min. = minimum; 25% and 75% = first and third quartiles; Max. = maximum; SD = standard deviation.

Greenness saturation is evident in Figure 2 below. A second-order polynomial regression of NDVI against PB across the entire dataset revealed a non-linear relationship with R^2 = 0.20 (Figure 2a), clearly demonstrating the saturation of NDVI at high biomass levels where the index plateaus while PB continues to increase, and progressively larger test splits (80:20, 70:30, 60:40; Figure 2b–d) retain similar slopes and R^2 values (0.20–0.25). The fan-shaped residual pattern shows that NDVI plateaus near 0.80 while PB continues to rise beyond 3000 kg DM ha^{-1}, with NDRE emerging as the sole significant positive coefficient (+3701 kg DM ha^{-1} unit^{-1}, $p < 0.001$).

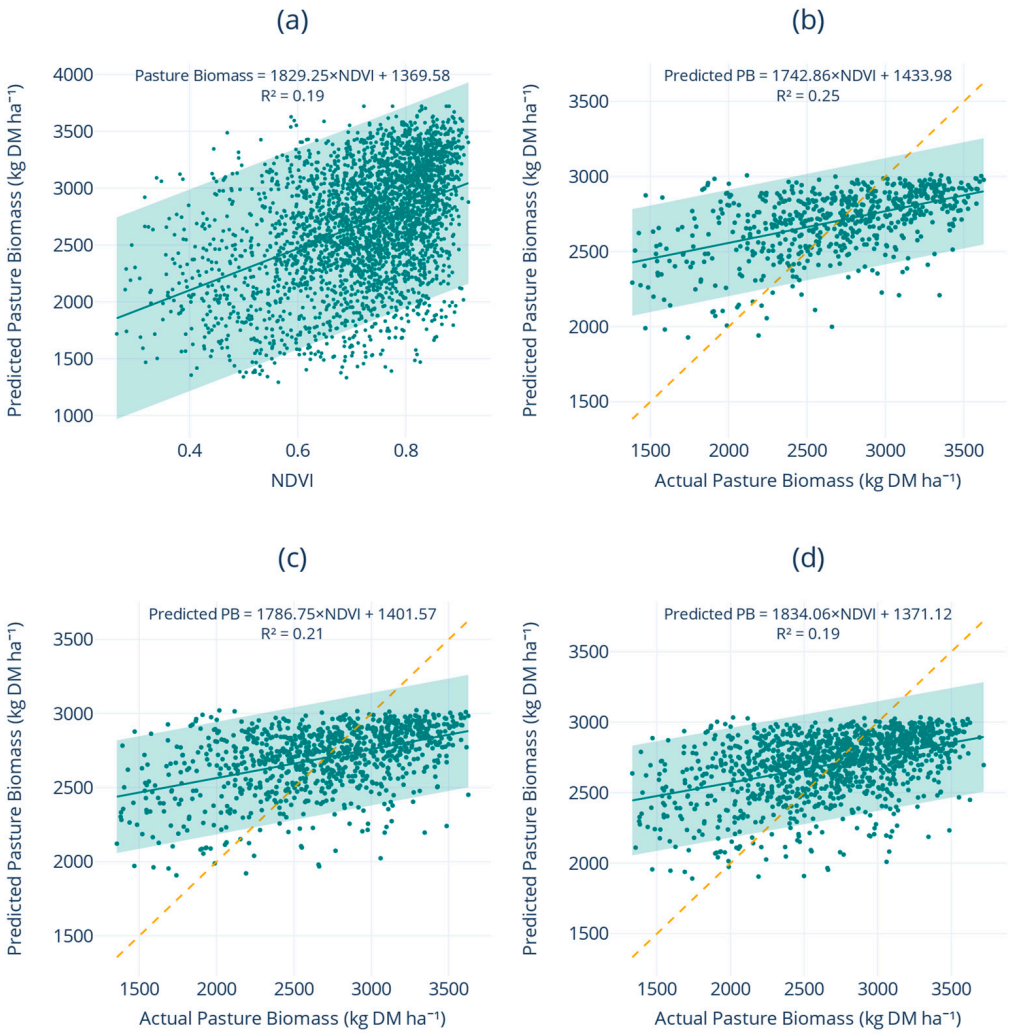


Figure 2. Combined regression and validation of pasture biomass (PB) estimates. Panel (**a**) shows the linear regression (LR) between Normalised Difference Vegetation Index (NDVI) and rising plate-meter biomass (kg DM ha^{-1}) measurements, with a teal regression line, a shaded 95% prediction interval, and the fitted equation plus R^2 annotation. Panels (**b**–**d**) depict actual versus predicted PB for 80:20, 70:30 and 60:40 train–test splits using LR, respectively; each includes a teal regression line, an orange dashed 1:1 line, the 95% confidence band around the predicted values, and the corresponding R^2.

Inter-variable relationships, visualised in the Pearson-correlation heat map (Figure 3), partitioned predictors into three main clusters. First, the ten Sentinel-2 bands are highly auto-correlated ($r > 0.70$), a redundancy that later justifies dimensional reduction or tree-based feature selection. Second, the greenness–biomass cluster shows positive correlations between PB and composite indices, namely NDRE ($r = 0.49$), EVI (0.45), NDVI (0.44), and SAVI (0.44), and negative correlations with red and SWIR reflectance ($r -0.39$), reflecting the classical red–NIR contrast and water-absorption effects. Third, the weather cluster revealed tight coupling among maximum temperature, evapotranspiration, and solar radiation ($r \approx 0.75$), but only weak instantaneous ties to PB ($r \leq 0.18$). The five strongest positive and negative coefficients highlight the immediate value of red-edge indices, with NDRE and EVI as the most informative single predictors, whereas high-wavelength reflectance exerts the strongest damping influence.

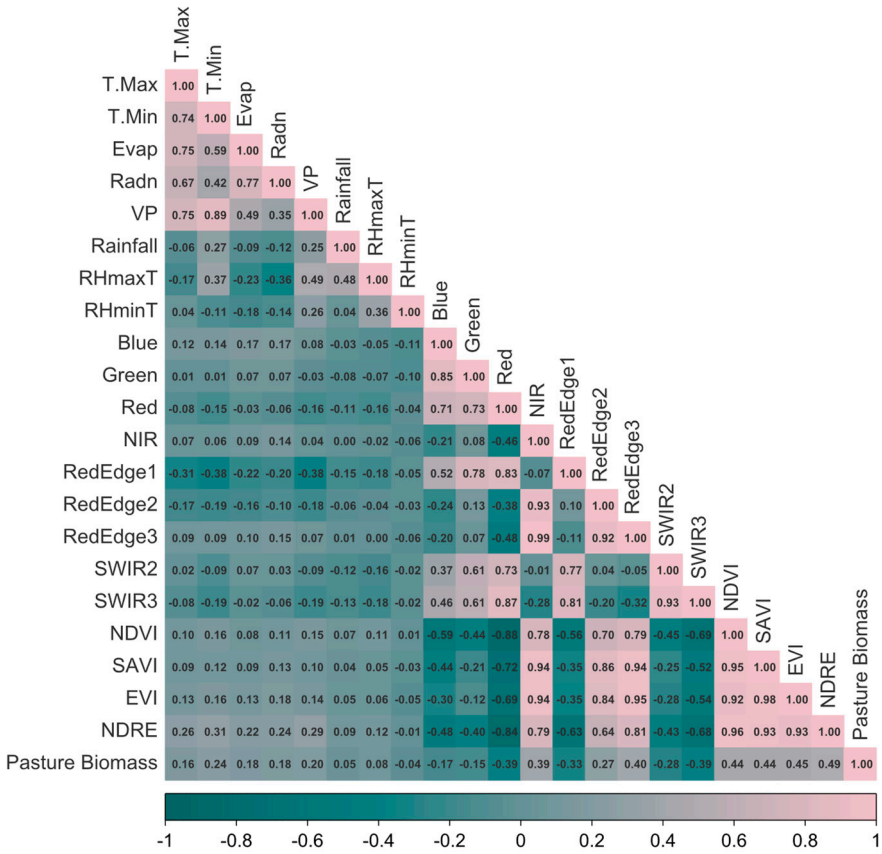


Figure 3. Pearson-correlation heatmap of selected biophysical, spectral, and meteorological variables. Pasture biomass (PB, kg DM ha^{-1}); weather variables include daily maximum temperature (T.Max, °C), daily minimum temperature (T.Min, °C), evapotranspiration (Evap, mm), incoming solar radiation (Radn, MJ m^{-2}), vapour pressure (VP, kPa), precipitation (Rainfall, mm), maximum relative humidity (RHmaxT, %), and minimum relative humidity (RHminT, %); spectral bands comprise Blue (blue band), Green (green band), Red (red band), NIR (near-infrared band), RedEdge1, RedEdge2, and RedEdge3 (red edge bands 1−3), and SWIR2 and SWIR3 (short-wave infrared bands 2−3); vegetation indices include NDVI (Normalised Difference Vegetation Index), SAVI (Soil Adjusted Vegetation Index), EVI (Enhanced Vegetation Index), and NDRE (Normalised Difference Red Edge Index).

3.2. Baseline Model Development and Predictive Accuracy

Eight regression algorithms were benchmarked on the non-interpolated dataset under four predictor configurations that varied the presence of weather variables and vegetation indices. Each pipeline applied one-hot encoding to categorical features, standard scaling to numeric features, and underwent hyper-parameter tuning via a five-fold, three-repeat Repeated K-Fold search that minimised MAE (Section 2.4.1). An 80:20 random split (n = 2528 training set, 633 test set) provided an unseen test set, while two independent hold-out samples supplied geographic and temporal validation.

Table 2 summarises test-set results for the configuration that proved most dependable comprising all Sentinel-2 spectral bands combined with concurrent weather predictors but excluding vegetation indices. Within this setting, XGBoost achieved the lowest test error of all pipelines (RMSE = 313 kg DM ha^{-1}, MAE = 243 kg DM ha^{-1}, R^2 = 0.63) and the highest cross-validated score (CV MAE = 246 kg DM ha^{-1}). To demonstrate the value of integrating full spectral information, XGBoost performance was compared against NDVI-based approaches: when trained using NDVI alone, performance was substantially lower (RMSE = 439 kg DM ha^{-1}, MAE = 359 kg DM ha^{-1}, R^2 = 0.28), while combining NDVI with weather factors improved results (RMSE = 363 kg DM ha^{-1}, MAE = 284 kg DM ha^{-1}, R^2 = 0.50) but still underperformed regarding the full spectral approach by 41 kg DM ha^{-1}

in MAE and 13 percentage points in R^2. Random forest and gradient boosting models followed closely (MAE ≈ 260 kg DM ha^{-1}, R^2 ≈ 0.60), whereas linear, k-nearest-neighbour, and support vector regressions lagged by 40–50 kg DM ha^{-1}. Stand-alone decision trees performed worst with MAE = 331 kg DM ha^{-1} and R^2 = 0.37, highlighting the stabilising value of ensemble averaging in a feature space dominated by collinear reflectance bands.

Table 2. Test-set and cross-validation performance of the eight regression algorithms all using Sentinel-2 bands and weather variables (no vegetation indices).

Model	CV Score	MSE	RMSE	MAE	R^2
Decision Tree	−336.38	168,853.56	410.92	331.35	0.37
Support Vector Regression	−289.99	138,196.95	371.75	289.81	0.48
Linear Regression	−284.00	129,687.39	360.12	285.11	0.51
Lasso Regression	−283.89	129,417.34	359.75	284.98	0.52
K-Nearest Neighbours	−283.06	126,959.97	356.31	281.89	0.52
Random Forest	−274.80	105,868.82	325.37	259.31	0.60
Gradient Boosting Machines	−249.08	105,539.33	324.87	252.70	0.60
Extreme Boosting Regressor	−246.41	98,036.41	313.11	243.12	0.63

CV Score is the negative mean absolute error (MAE) averaged across a five-fold, three-repeat cross-validation; MSE = mean squared error, RMSE = root mean squared error, and MAE (all in kg DM ha^{-1}); R^2 is the coefficient of determination on the independent 20% test split (n = 633). Extreme Boosting Regressor (XGBoost) records the lowest error metrics, underscoring the advantage of tree-ensemble methods for modelling PB from high-dimensional spectral and weather data.

Figure 4 shows the association between actual and predicted PB on the non-interpolated test set for four predictor scenarios. Including the four derived indices (NDVI, SAVI, EVI, NDRE) lowered the XGBoost test-set MAE marginally from 243 to 238 kg DM ha^{-1} and nudged R^2 to 0.65 (Figure 4b), yet this apparent gain disappeared on both validation sets. For the November 2024 paddocks, the index-rich model returned MAE = 315 kg DM ha^{-1} and R^2 = 0.27, whereas the index-free model achieved MAE = 296 kg DM ha^{-1} and R^2 = 0.33 (Figure 5a). A similar pattern emerged on the sixth-paddock sample (MAE = 264 versus 256 kg DM ha^{-1}, R^2 = 0.29 versus 0.35; Figure 5b).

Predictor sets that omitted either weather data or indices performed consistently worse (Figure 4a,c). Removing meteorological inputs lowered test-set R^2 by 5–6 points and raised MAE by roughly 30 kg DM ha^{-1}. Validation errors exceeded test errors by about 10%, and the ranking of model quality remained unchanged, with XGBoost ahead of GBM and RF, followed by linear methods. XGBoost presented test-set MAE 12% lower relative to linear regression and maintained that margin on both external validation datasets.

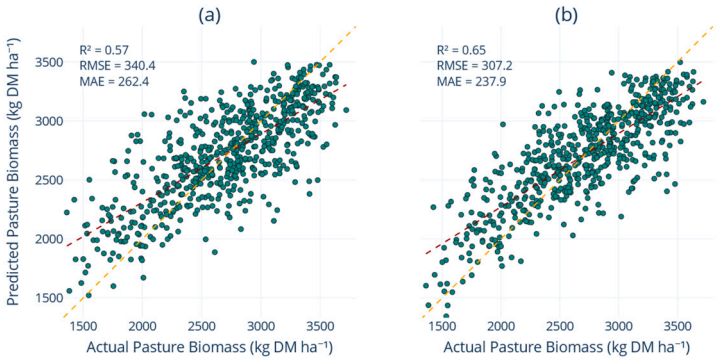


Figure 4. *Cont.*

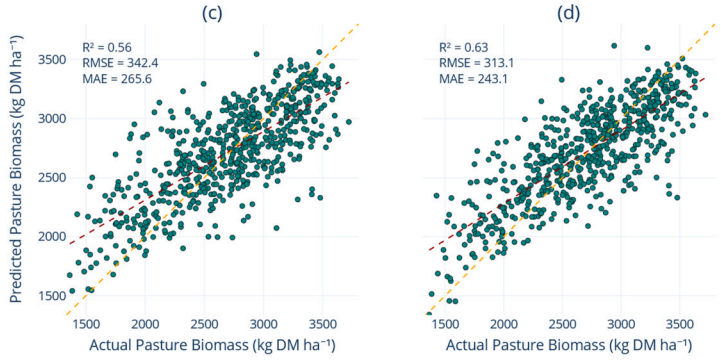


Figure 4. Scatterplots of actual versus predicted PB (kg DM ha^{-1}) on the non-interpolated test set (n = 633 rows) for four predictor scenarios. Panels (**a**–**d**) correspond, respectively, to (**a**) all spectral bands plus vegetation indices (NDVI, SAVI, EVI, NDRE) without weather data; (**b**) all spectral bands with weather variables and vegetation indices; (**c**) spectral bands alone without indices or weather data; and (**d**) spectral bands with weather variables but without indices. Each panel displays its R^2, RMSE, and MAE (kg DM ha^{-1}) in the top left corner. The maroon dashed line represents the fitted regression line, and the orange dashed line shows the one-to-one line.

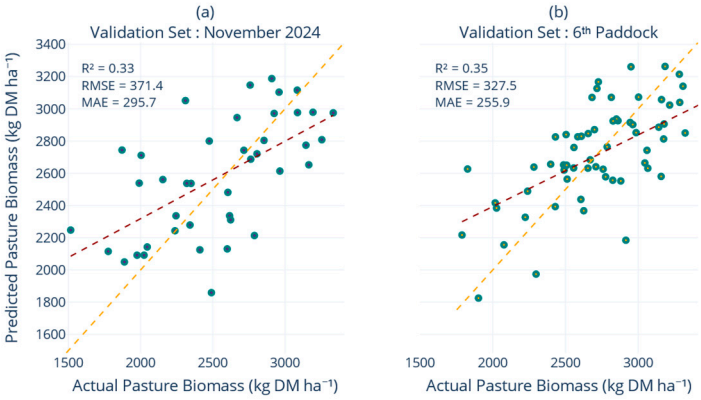


Figure 5. Scatterplots of actual versus predicted PB (kg DM ha^{-1}) on two independent validation sets for the non-interpolated scenario. All the bands with weather data without vegetation indices (NDVI, SAVI, EVI, NDRE). Panel (**a**) shows the November 2024 paddocks (the main five monitored paddocks per farm; n = 41 rows), and panel (**b**) shows the sixth extra paddock monitored for each farm (n = 63 rows) with axes starting at a minimum of 1500 kg DM ha^{-1}. Both subplots are based solely on spectral bands (no vegetation indices) and include their R^2, RMSE, and MAE (kg DM ha^{-1}) in the top-left corner. The maroon dashed line represents the regression line, and the orange dashed line indicates the one-to-one line.

3.3. Effects of Feature Engineering and Data Augmentation

Interpolation of field measurements using four techniques, second-order polynomial, Gaussian radial-basis function (rbf), multiquadric radial-basis function (mq), and minimum-curvature spline (mcg), were applied under three scenarios as shown in Figure 5: Year1Set only, Year2Set only, or the combined dataset. These methods were benchmarked against the non-interpolated (Non_ip) baseline to assess their impact on model performance. Each experiment utilised the "All Bands with Weather Data without Indices" feature set and the XGBoost pipeline detailed in Table 2. Performance was evaluated using MSE, RMSE, MAE, and R^2 on the 80:20 test split and two independent hold-out validation sets, as defined in Section 2.4.3. A comparative analysis of these methods is presented in Figure 5. The results revealed that mq interpolation produced the most substantial improvements on the 80:20 test split. Augmenting the training data with approximately 30 per cent synthetic mq-derived observations increased the test R^2 from 0.63 to 0.70 and reduced the MAE from 243 to 216 kg DM ha^{-1}, an 11% improvement over the baseline. A parallel evaluation using

the mq-interpolated dataset (Table 3) showed that data augmentation through interpolation improved test-set accuracy and raised cross-validated performance relative to the non-interpolated baseline reported in Table 2. As shown in Figure 6, significant gains were also apparent when interpolation was confined to Year 1 data alone; this approach lifted the R^2 to 0.71 and reduced the MAE to 198 kg DM ha^{-1}, achieving the highest cross-validated score.

Table 3. Test-set and cross-validation performance of the eight regression algorithms using the mq-interpolated dataset under the "All Bands with Weather Data without Indices" feature configuration.

Model	CV Score	MSE	RMSE	MAE	R^2
Decision Tree	−324.31	190,801.32	436.81	308.96	0.33
Support Vector Regression	−310.76	173,507.53	416.54	322.48	0.39
Linear Regression	−308.63	171,153.62	413.71	325.34	0.40
Lasso Regression	−308.63	171,145.21	413.70	325.34	0.40
K-Nearest Neighbours	−261.54	122,847.55	350.50	253.14	0.57
Random Forest	−246.65	100,377.68	316.82	237.21	0.65
Gradient Boosting Machines	−233.05	87,026.31	295.00	218.05	0.70
Extreme Boosting Regressor	−231.89	85,631.65	292.63	216.13	0.70

CV Score is the negative mean absolute error (MAE) averaged across a five-fold, three-repeat cross-validation; MSE = mean squared error, RMSE = root mean squared error, and MAE (all in kg DM ha^{-1}); R^2 is the coefficient of determination on the independent 20% interpolated test split (n = 982). Extreme Boosting Regressor (XGBoost) records the lowest error metrics, underscoring the advantage of tree-ensemble methods for modelling PB from high-dimensional spectral and weather data.

The robustness of the final multiquadric-augmented model was reflected on the external validation sets (see Section 2.4.3), with detailed results visualised in Figure 7. On the November 2024 paddocks, the model achieved an R^2 of 0.44, MSE of 100,489 kg^2 DM^2 ha^{-2}, RMSE of 317 kg DM ha^{-1}, and an MAE of 267 kg DM ha^{-1}, outperforming the non-interpolated baseline (R^2 = 0.33, MSE = 123,456 kg^2 DM^2 ha^{-2}, RMSE = 351 kg DM ha^{-1}, MAE = 296 kg DM ha^{-1}). The performance on the sixth-paddock sample was better, delivering an R^2 of 0.48, MSE of 78,225 kg^2 DM^2 ha^{-2}, RMSE of 280 kg DM ha^{-1}, and an MAE of 235 kg DM ha^{-1}, whereas other interpolation methods failed to surpass an R^2 of 0.37. As illustrated in Figure 6, predictions from the mq model show strong agreement between observed and predicted values for the test set and both validation samples, while maintaining homogeneous residual dispersion. The multiquadric augmentation increased the training data size by roughly 30 per cent, reduced the MSE by 20,000 to 25,000 kg^2 DM^2 ha^{-2}, reduced the RMSE by 30 to 40 kg DM ha^{-1}, reduced the MAE by 20 to 30 kg DM ha^{-1}, and raised the R^2 by 5 to 7 percentage points, while retaining predictive accuracy when applied to new seasons or paddocks unseen during model fitting.

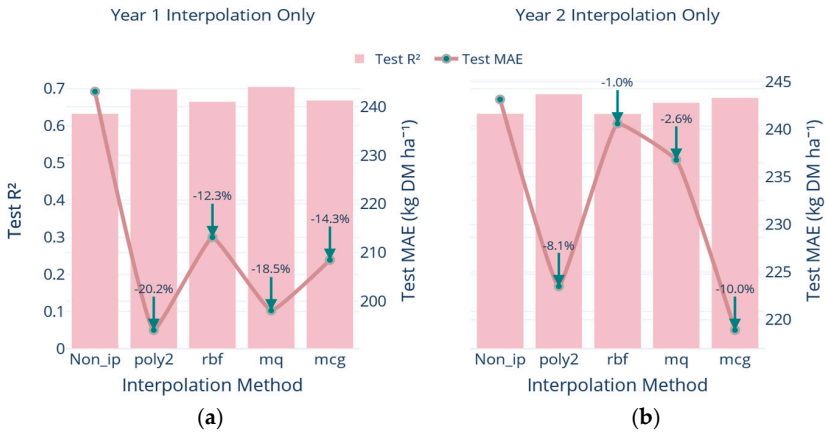
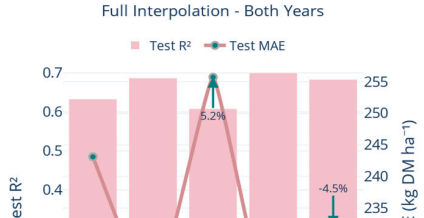
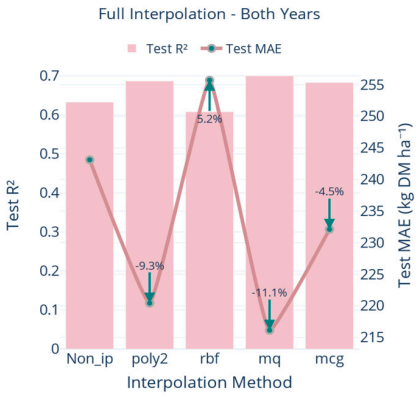


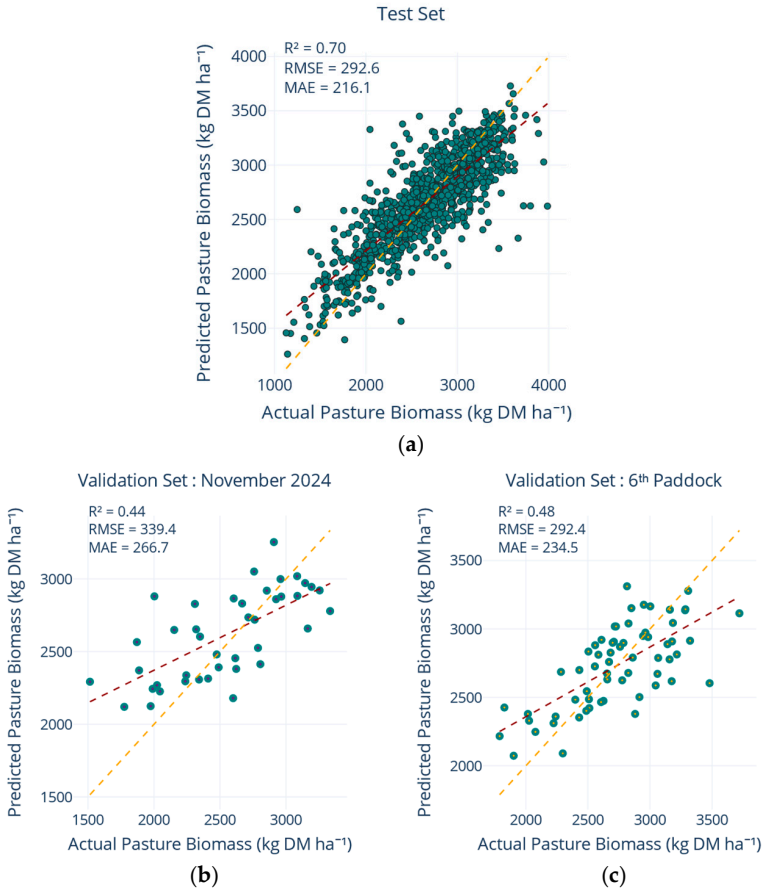
Figure 6. *Cont.*





(**c**)

Figure 6. Impact of different interpolation methods on XGBoost model performance using the "All the Bands with Weather Data without Indices" feature set. In every panel, test R^2 is shown as pink bars and test MAE by a light-pink line with teal markers; at each MAE point, the teal arrow indicates the percentage increase (upward) or decrease (downward) relative to the non-interpolated (Non_ip) baseline. Panel (**a**) (top left) shows Year1Set (pre −April 2023) interpolation methods, polynomial (poly2), RBF, multiquadric (mq), and minimum-curvature (mcg). Panel (**b**) (top right) shows the same four methods applied in Year2Set (post −1 April 2023). Panel (**c**) (centre bottom) illustrates the performance of these four interpolation methods when applied to the full two-year period.



(**a**)

(**b**) (**c**)

Figure 7. Observed versus predicted PB (kg DM ha^{-1}) obtained with the XGBoost model trained on the multiquadric interpolated data. Panels: (**a**) test set; (**b**) validation set—November 2024 (main five monitored paddocks per farm; n = 41); (**c**) validation set—sixth extra paddock per farm (n = 63). The maroon dashed line represents the regression fit, and the orange dashed line denotes the 1:1 relationship between actual and predicted PB estimates.

3.4. Temporal–Spatial Generalisation and Progressive Training

Training exclusively on Year 1 records and projecting onto Year 2 data provides a stringent test of temporal generalisation. When the model was fitted to the raw, non-interpolated Year1Set (1785 rows) and evaluated on the full Year2Set hold-out (1369 rows), it achieved R^2 = 0.24, MSE = 195,667 kg^2 DM^2 ha^{-2}, RMSE = 442 kg DM ha^{-1}, and MAE = 349 kg DM ha^{-1} (Figure 8a). However, replacing the raw Year1Set with multiquadric-interpolated values (4242 rows) reduced performance (R^2 = 0.14, MSE = 220,170 kg^2 DM^2 ha^{-2}, RMSE = 469.22 kg DM ha^{-1}, MAE = 368 kg DM ha^{-1}) as shown in (Figure 8b), while interpolating both years further decreased accuracy (R^2 = 0.10, MSE = 259,875 kg^2 DM^2 ha^{-2}, RMSE = 510 kg DM ha^{-1}, MAE = 397 kg DM ha^{-1}).

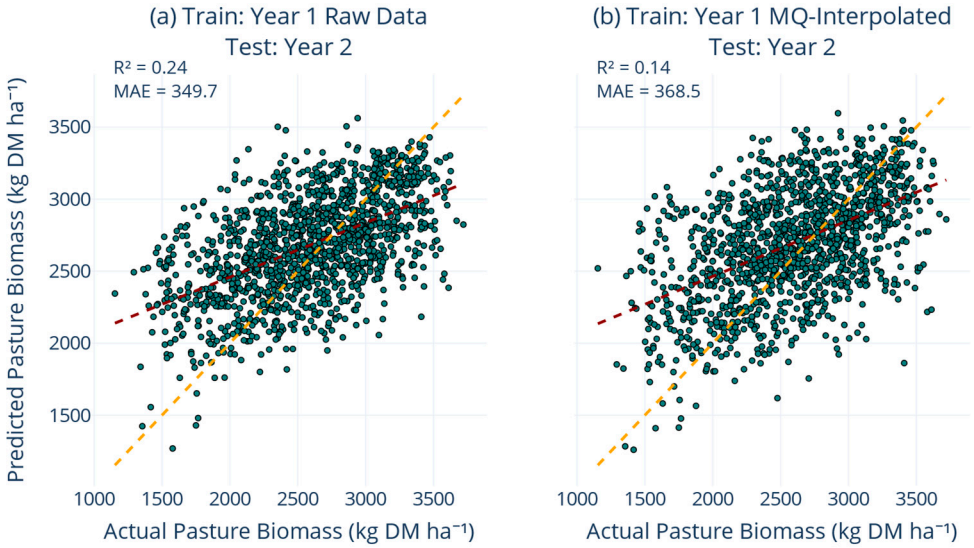


Figure 8. Comparison of observed versus predicted PB (kg DM ha^{-1}) when training on Year 1 data and testing on Year 2 data. Panel (**a**) shows the model trained on raw (non-interpolated) Year 1 data → Year 2 Test, while panel (**b**) shows the model trained on Year 1 multiquadric-interpolated data → Year 2 Test. Each panel plots actual versus predicted values for the Year 2 hold-out, with annotated R^2 and MAE. The maroon dashed line represents the regression fit, and the orange dashed line denotes the 1:1 relationship between actual and predicted PB estimates.

Weekly-observation-mask subsets captured one, two, three, or four weeks of each monthly cycle, corresponding to roughly 25%, 50%, 75%, and 100% of the available chronology. On the non-interpolated data, the test-set R^2 climbed steadily from 0.31 with the sparsest subset to 0.60 with the full dataset, while MAE fell from 328 to 244 kg DM ha^{-1} (Figure 9a). Validation on the November 2024 paddocks (n = 41) followed the same trajectory: R^2 improved from 0.10 to 0.32 and MAE declined from 357 to 307 kg DM ha^{-1}. Accuracy on the sixth-paddock set rose in parallel, reaching R^2 = 0.30 and MAE = 261 kg DM ha^{-1} for the complete training span.

The gains were larger when the progressively expanding archive also contained multiquadric interpolations. With one quarter of the year retained, the interpolated subset already matched the four-week raw model (R^2 = 0.44, MAE = 294 kg DM ha^{-1}). Assimilating the entire augmented chronology (4486 rows) lifted the test R^2 to 0.68 and pushed MAE down to 208 kg DM ha^{-1}, a 29% reduction relative to the 1 W baseline (Figure 9b). Crucially, these improvements carried into space and time: on the November 2024 validation R^2 reached 0.41 with MAE = 288 kg DM ha^{-1}, and on the sixth-paddock set R^2 climbed to 0.37 with MAE = 255 kg DM ha^{-1}. The scatterplot in Figure 9b shows strong agreement between observed and predicted values with a homogeneous spread of residuals, confirming that the progressive schedule prevents the interpolation surface from over-fitting transient

fluctuations. Spatial leave-farm-out experiments provided complementary insight into how well the interpolated model generalises across farms. Holding out one entire farm while training on the remaining 15 farms using the multiquadric-augmented dataset yielded R^2 = 0.46 and MAE = 299 kg DM ha^{-1}.

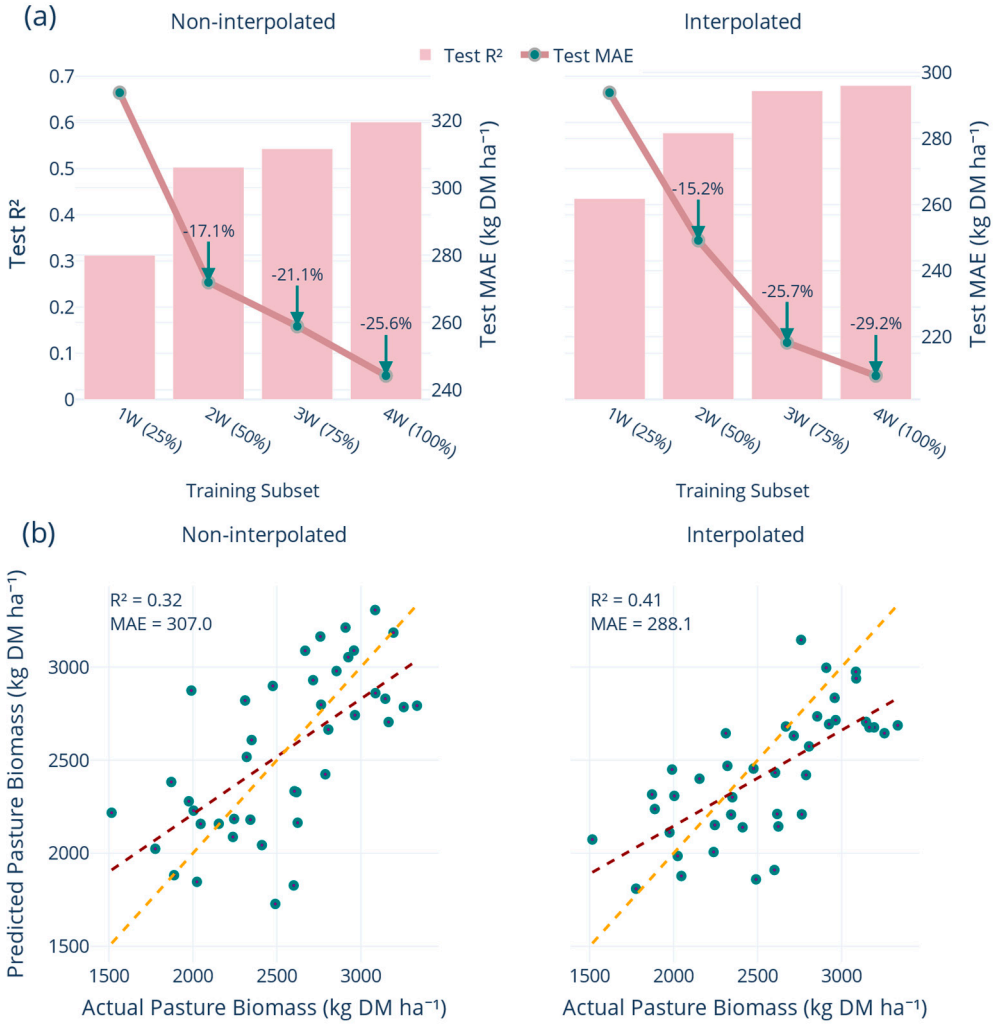


Figure 9. Progressive-training performance of the XGBoost PB model built from the "All the Bands with Weather Data without Indices" feature set. (**a**) Pink bars show the test-set R^2 for each weekly-observation-mask (WOM) subset: 1 W (25%), 2 W (50%), 3 W (75%), and 4 W (100%). The light-pink line with teal markers depicts the corresponding test MAE; teal arrows indicate the percentage change in MAE relative to the 1 W (25%) baseline (upward for increases, downward for decreases). (**b**) Observed versus predicted PB (kg DM ha^{-1}) on the validation set—November 2024 (main five monitored paddocks per farm; n = 41). The maroon dashed line represents the regression fit, and the orange dashed line denotes the 1:1 relationship between actual and predicted PB estimates.

A final benchmark evaluated the application of the multiquadric-augmented Sentinel-2 XGBoost model (Section 3.3) by comparing its estimates against Pasture.io; https://pasture.io (PIO), a commercial PB estimation platform. The comparison was conducted on 605 matched observations, where both the developed model and PIO provided estimates for identical paddocks and dates from the 20% test set. The agreement between the two approaches was strong, with an MAE of 240 kg DM ha^{-1} and the open-source model explaining 66% of the variance in PIO estimates (R^2 = 0.66). The key findings from this comprehensive analysis have been summarised in Table 4 below.

Table 4. Summary of key findings from PB estimation using integrated Sentinel-2 and ML approaches used in this study.

Analysis Component	Key Finding	Performance Metrics
Vegetation Index Limitations	NDVI saturates at ~0.80 when PB exceeds 3000 kg DM ha^{-1}	R^2 = 0.19, fan-shaped residual pattern
Optimal Model Configuration	XGBoost with full spectral bands + weather data without indices (non-interpolated)	R^2 = 0.63, MAE = 243 kg DM ha^{-1}
Data Augmentation Impact	Multiquadric interpolation (~30% synthetic observations)	Improved to R^2 = 0.70, MAE = 216 kg DM ha^{-1}
External Validation	Interpolated model on independent validation sets	Validation Set1: R^2 = 0.44, MAE = 267 kg DM ha^{-1}; Validation Set2: R^2 = 0.48, MAE = 235 kg DM ha^{-1}
Progressive Training	Continuous updating (1–4 weeks training subsets)	29% MAE reduction from 1 W to 4 W (328 to 208 kg DM ha^{-1})
Spatial Generalisation	Leave-farm-out validation (trained on 15 farms, tested on 1)	R^2 = 0.46, MAE = 299 kg DM ha^{-1}
Commercial Comparison	Benchmarked against Pasture.io platform	R^2 = 0.66, MAE = 240 kg DM ha^{-1}
Temporal Limitations	Year1Set to Year2Set extrapolation challenges	Performance declined when training on single year only

4. Discussion

4.1. Exploratory Analysis and Vegetation Index Limitations

Descriptive analysis revealed that PB is governed by seasonality and regional setting, with composite spectral indices explaining daily variation more effectively than instantaneous weather readings. The Pearson correlation heat map (Figure 3) partitioned predictors into three main clusters that informed the modelling strategy. The high auto-correlation among the ten Sentinel-2 bands ($r > 0.70$) justified the use of tree-based feature selection methods rather than attempting manual dimensional reduction. The greenness-biomass cluster showed that while NDRE ($r = 0.49$), EVI (0.45), NDVI (0.44) and SAVI (0.44) all correlated positively with PB, none achieved correlation coefficients above 0.50, indicating substantial unexplained variance. The weather cluster revealed tight coupling among maximum temperature, evapotranspiration and solar radiation ($r \approx 0.75$) but only weak instantaneous ties to PB ($r \leq 0.18$), emphasising the lagged influence of weather on pasture growth rather than same-day effects. These insights guided the development of multi-feature ML pipelines that merged meteorological variables, full-band Sentinel-2 reflectance, red-edge-enhanced indices and categorical predictors to capture the residual variability in PB. The univariate regression analysis (Figure 2) demonstrated that increased sampling does not alleviate the saturation problem inherent in vegetation indices. Saturation occurs when the relationship between a vegetation index and biomass becomes non-linear and eventually plateaus, meaning that increases in biomass no longer result in proportional increases in the index value. The fan-shaped residual pattern confirms that NDVI plateaus near 0.80 while PB continues to rise beyond 3000 kg DM ha^{-1}. Even red-edge information (NDRE), which emerged as the sole significant positive coefficient (+3701 kg DM ha^{-1} unit^{-1}, $p < 0.001$), cannot fully capture PB variability once canopies exceed approximately 2800 kg DM ha^{-1}. This saturation fundamentally limits the utility of univariate spectral regressions at high canopy density, where accurate PB estimation is most critical for practical farm management.

An important consideration is that PB variability reflects not only environmental conditions but also management practices, particularly grazing pressure, which can rapidly

alter available biomass independent of weather or spectral signals [39,40]. In the absence of explicit management variables, environmental predictors may partly function as categorical site indicators that implicitly capture farm-specific utilisation pattern. This limitation reinforces the necessity of frequent ground-truth calibration through RPM measurements, which inherently integrate both environmental and management effects at each sampling occasion [9,10]. The fortnightly to weekly sampling regime employed here was specifically designed to capture these combined dynamics, while meteorological variables enhanced temporal interpolation between observations [41].

4.2. Overcoming Vegetation Index Saturation Through Multi-Spectral Integration

The primary challenge addressed in this study was the saturation of traditional vegetation indices like NDVI under dense canopies, which typically occurs when PB exceeds approximately 3 tonnes DM ha^{-1} [42]. This saturation fundamentally limits the utility of conventional remote sensing approaches for practical farm management, where accurate PB estimation is most critical at higher PB levels. The univariate regression analysis (Figure 2) clearly illustrated this limitation, with NDVI plateauing near 0.80 whilst PB continued to rise beyond 3000 kg DM ha^{-1}, resulting in the characteristic fan-shaped residual pattern that confirms the inadequacy of NDVI alone for explaining PB variability.

The integrated approach of combining raw Sentinel-2 reflectance with weather variables demonstrates a clear pathway beyond these limitations. The progression from NDVI-only (R^2 = 0.28, MAE = 359 kg DM ha^{-1}) to NDVI plus weather factors (R^2 = 0.50, MAE = 284 kg DM ha^{-1}) to full spectral bands with weather data (R^2 = 0.63, MAE = 243 kg DM ha^{-1}) clearly shows that whilst weather data provides valuable orthogonal information, the complete Sentinel-2 spectral suite captures essential biophysical information that traditional indices cannot adequately represent. This finding aligns with recent research by Jennewein, et al. [43], who achieved R^2 = 0.70 for crop biomass/PB estimation using multi-sensor proximal remote sensing combining multiple satellite platforms. However, our current study demonstrates that comparable performance can be achieved using only freely available Sentinel-2 data, highlighting the cost-effectiveness and accessibility of this approach for widespread adoption across diverse farming systems. Collectively these findings demonstrate that PB is governed by a variety of factors such as seasonality and regional setting, that composite spectral indices explain daily variation more effectively than instantaneous weather readings, and that greenness saturation limits the usefulness of univariate spectral regressions at high canopy density.

Guerini Filho, et al. [44] showed that Sentinel-2 imagery combined with vegetation indices could predict natural grassland PB with R^2 values ranging from 0.51 to 0.65. However, our current study demonstrates that robust predictive performance can be achieved even in the absence of explicit vegetation indices, indicating that raw reflectance bands, especially in the red-edge and short-wave infrared regions, inherently contain the necessary biophysical information. This finding is further supported by Gargiulo, et al. [18], who reported R^2 = 0.72, RMSE = 255 kg DM/ha^{-1} when combining Sentinel-2 with Planet CubeSats data, but required multiple commercial satellite sources. The decision to exclude vegetation indices from the final model aligns with findings from Morse-McNabb, et al. [19], who demonstrated that including SWIR bands substantially enhanced yield prediction accuracy with Sentinel-2 when predicting PB above 3000 kg DM ha^{-1}, improving R^2 from 0.79 to 0.90 and reducing RMSE by nearly 200 kg DM ha^{-1}. Ogungbuyi, et al. [45] similarly highlighted the limitations of index-only approaches, noting that despite a moderate correlation with total PB (R^2 = 0.43), the associated MAE of 871.83 kg DM ha^{-1} was too large for practical application. In contrast, the inclusion of SWIR bands, as demonstrated in the current study, not only minimised preprocessing requirements but also improved

generalisation on external datasets. The validation results (Figure 5) confirmed that raw red-edge and short-wave-infra-red bands already carry the information encapsulated by the indices, with explicit index calculation adding noise without improving generalisation.

Although weather showed weak instantaneous correlations with PB (Section 3.1), concurrent weather still provides orthogonal information that improves prediction. Removing meteorological inputs lowered test-set R^2 by 5-6 points and raised MAE by roughly 30 kg DM ha^{-1}, confirming that weather variables capture aspects of pasture condition not reflected in same-day spectral measurements. The ranking of model quality remained unchanged across validation sets, with XGBoost ahead of gradient boosting and random forest, followed by linear methods, showing that ensemble tree algorithms capture non-linear interactions between weather and reflectance that translate beyond the calibration domain. Stand-alone decision trees performed worst, highlighting the stabilising value of ensemble averaging in a feature space dominated by collinear reflectance bands.

Consistent with our findings, Chen, et al. [46] reported optimal PB prediction when all Sentinel-2 bands, NDVI, and weather variables were included, yielding R^2 approximately 0.60 and MAE approximately 262 kg DM ha^{-1}. This study matched these benchmarks without requiring explicit inclusion of vegetation indices, reinforcing the argument that full spectrum inputs provide a richer predictive foundation than derived indices alone. In a broader modelling context, Netsianda and Mhangara [4] combined Sentinel-2 bands, NDVI, and elevation data to estimate PB using RF and GBM algorithms, achieving an R^2 of 0.73. Whilst their approach involved multiple data modalities, our findings demonstrate comparable or superior results using only satellite and weather data, without requiring additional topographic inputs.

While this study demonstrated that XGBoost effectively captured information from raw SWIR2 and SWIR3 bands through non-linear feature interactions, future research could explore whether explicit SWIR-based indices such as the Cellulose Absorption Index (CAI) or Normalised Difference Lignin Index (NDLI) provide additional interpretability or improve performance in simpler, more interpretable models (e.g., linear regression or decision trees) that may be preferred in some operational contexts [47]. Recent studies have demonstrated that SWIR bands are particularly effective for predicting high pasture biomass (>4000 kg DM ha^{-1}) where chlorophyll-based indices like NDVI saturate [19,48], and that explicit SWIR-enhanced indices can provide clearer mechanistic insights into canopy structural properties [49,50]. However, our results confirm that ensemble tree-based models can effectively learn these relationships directly from raw spectral bands without requiring pre-calculated indices.

4.3. Data Augmentation Through Multiquadric Interpolation

Cloud cover and satellite revisit cycles create inevitable temporal gaps in optical remote sensing data, a fundamental limitation recognised across the remote sensing literature [4,16]. The multiquadric radial basis interpolation approach employed in this study addresses this limitation by filling gaps in the RPM time series, creating interpolated observations based on actual observations that improve model training without introducing unrealistic temporal assumptions. The interpolation process accounts for pasture growth between measurements by fitting a smooth surface through the observed RPM points, with the multiquadric function providing a mathematically principled way to estimate intermediate values that reflect the gradual accumulation and depletion of biomass between actual field measurements. By confining interpolation strictly to the temporal span defined by observed measurements and requiring at least three actual observations per paddock, the method avoids extrapolation whilst capturing the underlying growth trajectory.

Whilst radial basis interpolation has found application in environmental and medical sciences, and some studies have explored interpolation methods for satellite data in agricultural applications, the specific application of multiquadric interpolation to bridge temporal gaps between ground-truth RPM measurements and cloud-affected satellite imagery in pasture biomass estimation represents a methodological contribution to this field. The observed improvement aligns with recent advances in agricultural data augmentation reported by Gracia Moisés, et al. [51], who demonstrated substantial error reductions using similar techniques in optical spectroscopy applications. However, the application to temporal gap-filling in satellite-ground data integration represents a distinct methodological contribution.

The multiquadric surface employed in this study demonstrated a favourable balance of speed, stability, and predictive effectiveness compared to alternative interpolation methods (Figure 6). This approach contrasted with other interpolation methods such as Gaussian process or Kriging augmentation, which, despite improving spatial homogeneity, are computationally intensive for national-scale datasets [52]. Recent mathematical advancements that generalise the multiquadric kernel for quasi-interpolation [53–55] suggest that even higher fidelity could be achieved as these theoretical formulations become operational in geospatial libraries. The trend-assisted multiquadric gridding approach demonstrated efficiency gains noted in Earth observation applications [31], supporting its adoption for operational pasture monitoring systems.

4.4. Progressive Training and Temporal Generalisation

The integration of interpolated rows within the progressive training regime demonstrated the practical benefit of continually updating the model. This trajectory corroborates the findings of Correa-Luna, et al. [10] that even a relatively small proportion of fortnightly ground observations, roughly 10% of paddock days, is sufficient to stabilise pasture predictions whilst emphasising the necessity of retraining the model on a comparable rolling schedule.

However, the temporal validation experiments (Figure 8) revealed important limitations of interpolation when applied to strict temporal extrapolation scenarios. Training exclusively on Year1Set and testing on Year2Set demonstrated that interpolation actually reduced performance compared to the raw data baseline (R^2 slipped from 0.24 to 0.14, MAE rose from 349 to 368 kg DM ha^{-1}). This apparent contradiction with the overall benefits of interpolation highlights the context-dependent nature of data augmentation techniques. This finding aligns with domain adaptation challenges widely reported in ML literature Meyer, et al. [56]. To understand this performance decline, we decomposed the prediction error into bias and variance components. The interpolated Year1Set model exhibited higher bias (systematic underestimation of Year2 biomass by approximately 15%) and similar variance compared to the raw model, suggesting that the interpolation surface captured Year1-specific patterns that did not transfer to Year2 conditions.

The observations generated through interpolation introduced bias when the training and testing data came from fundamentally different temporal distributions, as the interpolation surface fitted to Year 1 data encoded seasonal patterns specific to that year. However, when training and testing data came from similar temporal distributions (the 80:20 split from the combined dataset), interpolation significantly improved performance by providing a denser, more representative sample of the underlying growth patterns. In summary, these findings also indicate that the combined application of interpolation and progressive training effectively bridges inherent observational gaps associated with optical remote sensing and establishes a data-efficient method for maintaining accuracy as seasonal conditions change. The comprehensive strategy directly addresses model validity

and reliability across varied climatic, soil, and management contexts, thereby offering a robust and generalisable solution.

4.5. Model Performance and Validation Across Multiple Scales

The robustness of the final multiquadric-augmented model was reflected across external validation sets (Figure 7). The model achieved consistent performance on the November 2024 paddocks and the sixth-paddock validation sets, outperforming the non-interpolated baseline whilst other interpolation methods failed to surpass an R^2 of 0.37. Spatial leave-farm-out experiments confirmed that a PB model trained on one set of paddocks inevitably experiences a reduction in accuracy when applied to entirely excluded fields. In our case, the non-interpolated version surrendered roughly one quarter of its explanatory power under this scenario, a decline consistent with cross-location deterioration reported by Smith, et al. [57]. However, the introduction of multiquadric interpolation significantly mitigated this loss, helping the model maintain a more balanced representation of the growth envelope and maintaining the MAE on the held-out farms below 300 kg DM ha^{-1}, while explaining nearly half of the PB variance, surpassing the more complex ensembles assessed by Smith, et al. [57] at comparable extrapolation distances.

The disparity between robust performance on the internal 20% test split and comparatively weaker scores on independent validation sets elucidates the inherent challenges associated with model transferability, as widely reported in remote sensing applications where models are tested beyond their calibration domain. Validation errors exceeded test errors by about 10%, reflecting the additional temporal and geographic distance embodied in the hold-out samples. Nevertheless, the crucial need for regular retraining is emphasised by the observation that model accuracy declined on paddocks and seasons withheld from calibration, even with interpolation. Overall, the results demonstrate that temporal generalisation hinges on two complementary strategies, retaining the full historical record so that the model learns from experiencing the complete seasonal duration, and supplementing that record with cautiously smoothed interpolations that densify the signal without inflating noise. When applied together, these measures raise R2 by about 30% points and cut MAE by roughly 85 kg DM ha^{-1} compared with a model trained on a single season and maintain consistent accuracy when entire farms are withheld during testing. The consistency of these results across scenarios shows that the model, built with the full predictor set, extends reliably to farms unseen during calibration, reinforcing the temporal gains reported above.

4.6. Comparison with Commercial Decision-Support Systems

The comparison with PIO, a commercial PB estimation platform, provides crucial context for the application of research-based approaches. The strong agreement between the developed open-source model and the proprietary platform (R^2 = 0.66, MAE = 240 kg DM ha^{-1}) demonstrates that transparent, academic approaches can achieve comparable performance to commercial solutions across matched observations spanning multiple farms and production years [10]. This benchmark was performed across multiple farms and distinct production years, demonstrating that an openly documented and fully reproducible approach can achieve relatively high performance whilst retaining adaptability to integrate new sensors and management variables as they become available. Commercial platforms typically have access to multiple satellite sources, sophisticated atmospheric correction algorithms, and proprietary data fusion techniques, making this performance parity particularly significant. The comparison validates methodological choices whilst highlighting the potential for democratising precision agriculture technologies. Many existing commercial solutions require significant capital investment or ongoing subscription

costs that may be prohibitive for smaller farming operations, particularly in developing regions where cost-effective monitoring solutions are most needed.

4.7. Limitations and Future Research Directions

Several factors constrain the current implementation and highlight avenues for future research. The reliance on multiquadric interpolation presumes gradual temporal evolution of PB, a common approach for filling cloudy gaps in vegetation monitoring but one that risks obscuring sharp declines associated with grazing or drought. Combining optical-derived indices with sensors capable of detecting rapid canopy structural shifts, such as C-band radar backscatter, could offset this limitation [58]. Pasture greenness measures displayed the well-recognised saturation effect once biomass exceeded roughly 3000 kg DM ha^{-1}, thereby limiting model sensitivity in the critical range where tactical management adjustments are most often needed [46,59]. Acquiring richer spectral information, such as narrow-band hyperspectral or chlorophyll fluorescence signals, may offer enhanced sensitivity under high biomass and address this ceiling effect [13,60,61].

Additionally, weather predictors were restricted to daily aggregates, limiting capacity to reflect sub-daily factors such as vapour pressure deficit, refined soil moisture estimates, or high-resolution temperature extremes that could potentially capture short-lived stress events influencing pasture growth between satellite overpasses. The study spanned three districts within a single coastal climatic zone, and explicit testing of site-specific differences in management such as fertiliser timing or detailed pasture quality was not conducted. As noted by Holzworth, et al. [62], expansion of ground data across diverse weather and management systems is essential to evaluate broad-scale model transferability. The Year 1 to Year 2 transfer experiment reinforces this point because despite a threefold increase in sample size through interpolation, a model trained exclusively on the first season could not accommodate phenological shifts of the following year, with acceptable performance only restored after expanding the calibration base and retraining on multi-year data [63].

The demonstrated performance achievements using freely available satellite data have important implications for democratising precision agriculture technologies. The open-source approach provides full methodological transparency allowing for local adaptation and improvement, eliminates ongoing data acquisition costs through use of freely available Sentinel-2 data, and enables integration of additional sensor types or management variables as they become available. However, the current implementation may require some technical expertise, making it more immediately accessible to researchers, government agents, or technical staff rather than directly to individual farmers. Strategically, the progressive training paradigm establishes a pathway for continuously improving models that remain reliable as satellite technology, weather conditions, and management practices evolve.

Methodologically, this study significantly advances pasture remote sensing by demonstrating that integrating raw spectral, meteorological, and management data within a unified machine learning framework provides superior performance compared to models driven solely by vegetation indices. From a practical standpoint, the interpolation-enhanced model delivers near real-time PB estimates that align well with commercial platforms across diverse environments, offering producers a transparent and customisable alternative to proprietary solutions. Future work should focus on exploring ensemble combinations of optical and radar imagery, automating paddock-level recalibration using farmer-supplied RPM pasture biomass measurements, and integrating the model into grazing allocation tools to directly translate predictive accuracy into measurable productivity gains and farm profitability. These enhancements will enable integration into grazing allocation tools, establishing a dynamic, self-improving decision-support system that translates predictive accuracy into improved pasture utilisation and increased farm profitability.

5. Conclusions

This study demonstrates that accurate pasture biomass estimation can be achieved even where traditional vegetation indices saturate and Sentinel-2 imagery is intermittently obscured by cloud cover. By integrating raw multispectral reflectance, rising plate-meter measurements, daily weather data, and paddock-level metadata within a machine learning framework, the model effectively bridged the gap between satellite observations and ground measurements. The unified strategy prioritising full-band reflectance over vegetation indices achieved robust baseline performance (R^2 = 0.63, MAE = 243 kg DM ha^{-1}), substantially outperforming NDVI-based approaches. Multiquadric radial basis interpolation of field measurements addressed temporal gaps from cloud obstruction, augmenting the dataset with approximately 30% interpolated observations and improving performance to R^2 = 0.70 and MAE = 216 kg DM ha^{-1}. These improvements were observed across independent validation sets, with the November 2024 validation set achieving R^2 = 0.44 and MAE = 267 kg DM ha^{-1}, and the sixth-paddock validation set achieving R^2 = 0.48 and MAE = 235 kg DM ha^{-1}.

The implementation of progressive training with seasonally aligned observations-maintained model accuracy across temporal and spatial contexts. Leave-farm-out validation confirmed robust generalisation to unseen farms (R^2 = 0.46, MAE < 300 kg DM ha^{-1}), whilst comparison with the commercial Pasture.io platform demonstrated comparable performance using freely available data. Despite these advancements, several limitations warrant future research, as multiquadric interpolation may smooth abrupt pasture biomass declines following intensive grazing, and daily weather aggregates may overlook short-lived environmental events. Future work should focus on integrating finer-resolution sensors, developing automated paddock-level recalibration systems, and incorporating the model into grazing allocation tools to establish a dynamic decision-support system that translates predictive accuracy into improved pasture utilisation and farm profitability.

Author Contributions: Conceptualization, B.N.A., A.C., M.C.-L., C.E.F.C. and S.C.G.; methodology, B.N.A., A.C. and S.C.G.; software, B.N.A.; validation, B.N.A., S.C.G. and M.C.-L.; formal analysis, B.N.A.; investigation, B.N.A.; resources, B.N.A.; data curation, B.N.A.; writing—original draft preparation, B.N.A.; writing—review and editing, B.N.A., A.C., M.C.-L., C.E.F.C. and S.C.G.; visualization, B.N.A.; supervision, M.C.-L., C.E.F.C. and S.C.G.; project administration, S.C.G.; funding acquisition, S.C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dairy UP program, a collaborative RD&E program for New South Wales, Australia (www.dairyup.com.au) through the academic scholarship awarded to Blessing N. Azubuike, PI: Sergio C. Garcia.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This research was supported by the research scholarship granted by the DairyUP project at the University of Sydney.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3254–3264. [CrossRef]
2. Shi, Y.; Gao, J.; Li, X.; Li, J.; Torre, D.M.G.D.; Brierley, G.J. Improved Estimation of Aboveground Biomass of Disturbed Grassland through Including Bare Ground and Grazing Intensity. *Remote Sens.* **2021**, *13*, 2105. [CrossRef]
3. Pang, H.; Zhang, A.; Kang, X.; He, N.; Dong, G. Estimation of the Grassland Aboveground Biomass of the Inner Mongolia Plateau Using the Simulated Spectra of Sentinel-2 Images. *Remote Sens.* **2020**, *12*, 4155. [CrossRef]

4. Netsianda, A.; Mhangara, P. Aboveground biomass estimation in a grassland ecosystem using Sentinel-2 satellite imagery and machine learning algorithms. *Environ. Monit. Assess.* **2025**, *197*, 138. [CrossRef] [PubMed]
5. Wang, Z.; Ma, Y.; Zhang, Y.; Shang, J. Review of Remote Sensing Applications in Grassland Monitoring. *Remote Sens.* **2022**, *14*, 2903. [CrossRef]
6. Radočaj, D.; Šiljeg, A.; Marinović, R.; Jurišić, M. State of Major Vegetation Indices in Precision Agriculture Studies Indexed in Web of Science: A Review. *Agriculture* **2023**, *13*, 707. [CrossRef]
7. de Alckmin, G.T.; Kooistra, L.; Rawnsley, R.; Lucieer, A. Comparing methods to estimate perennial ryegrass biomass: Canopy height and spectral vegetation indices. *Precis. Agric.* **2021**, *22*, 205–225. [CrossRef]
8. Vidican, R.; Mălinaș, A.; Ranta, O.; Moldovan, C.; Marian, O.; Ghețe, A.; Ghișe, C.R.; Popovici, F.; Cătunescu, G.M. Using Remote Sensing Vegetation Indices for the Discrimination and Monitoring of Agricultural Crops: A Critical Review. *Agronomy* **2023**, *13*, 3040. [CrossRef]
9. Gargiulo, J.I.; Lyons, N.A.; Masia, F.; Beale, P.; Insua, J.R.; Correa-Luna, M.; Garcia, S.C. Comparison of Ground-Based, Unmanned Aerial Vehicles and Satellite Remote Sensing Technologies for Monitoring Pasture Biomass on Dairy Farms. *Remote Sens.* **2023**, *15*, 2752. [CrossRef]
10. Correa-Luna, M.; Gargiulo, J.; Beale, P.; Deane, D.; Leonard, J.; Hack, J.; Geldof, Z.; Wilson, C.; Garcia, S. Accounting for minimum data required to train a machine learning model to accurately monitor Australian dairy pastures using remote sensing. *Sci. Rep.* **2024**, *14*, 16927. [CrossRef]
11. Askari, M.S.; McCarthy, T.; Magee, A.; Murphy, D.J. Evaluation of Grass Quality under Different Soil Management Scenarios Using Remote Sensing Techniques. *Remote Sens.* **2019**, *11*, 1835. [CrossRef]
12. De Rosa, D.; Basso, B.; Fasiolo, M.; Friedl, J.; Fulkerson, B.; Grace, P.R.; Rowlings, D.W. Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Comput. Electron. Agric.* **2021**, *180*, 105880. [CrossRef]
13. Zhang, H.; Sun, Y.; Chang, L.; Qin, Y.; Chen, J.; Qin, Y.; Du, J.; Yi, S.; Wang, Y. Estimation of Grassland Canopy Height and Aboveground Biomass at the Quadrat Scale Using Unmanned Aerial Vehicle. *Remote Sens.* **2018**, *10*, 851. [CrossRef]
14. Naidoo, L.; van Deventer, H.; Ramoelo, A.; Mathieu, R.; Nondlazi, B.; Gangat, R. Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 118–129. [CrossRef]
15. Nguyen, H.T.T.; Doan, T.M.; Tomppo, E.; McRoberts, R.E. Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam. *Remote Sens.* **2020**, *12*, 1367. [CrossRef]
16. Dubovik, O.; Schuster, G.L.; Xu, F.; Hu, Y.; Bösch, H.; Landgraf, J.; Li, Z. Grand Challenges in Satellite Remote Sensing. *Front. Remote Sens.* **2021**, *2*, 619818. [CrossRef]
17. César, I.A.-M.; Guzman, D.; Casas, J.; Bastidas, M.; Polanco, J.; Valencia-Ortiz, M.; Montenegro, F.; Arango, J.; Ishitani, M.; Selvaraj, M.G. Predictive Modeling of Above-Ground Biomass in Brachiaria Pastures from Satellite and UAV Imagery Using Machine Learning Approaches. *Remote Sens.* **2022**, *14*, 5870. [CrossRef]
18. Gargiulo, J.; Clark, C.; Lyons, N.; de Veyrac, G.; Beale, P.; Garcia, S. Spatial and Temporal Pasture Biomass Estimation Integrating Electronic Plate Meter, Planet CubeSats and Sentinel-2 Satellite Data. *Remote Sens.* **2020**, *12*, 3222. [CrossRef]
19. Morse-McNabb, E.M.; Hasan, M.F.; Karunaratne, S. A Multi-Variable Sentinel-2 Random Forest Machine Learning Model Approach to Predicting Perennial Ryegrass Biomass in Commercial Dairy Farms in Southeast Australia. *Remote Sens.* **2023**, *15*, 2915. [CrossRef]
20. Amarsaikhan, E.; Erdenebaatar, N.; Amarsaikhan, D.; Otgonbayar, M.; Bayaraa, B. Estimation and mapping of pasture biomass in Mongolia using machine learning methods. *Geocarto Int.* **2023**, *38*, 2195824. [CrossRef]
21. Banerjee, P. MODIS-FIRMS and ground-truthing-based wildfire likelihood mapping of Sikkim Himalaya using machine learning algorithms. *Nat. Hazards* **2022**, *110*, 899–935. [CrossRef]
22. Vahidi, M.; Shafian, S.; Thomas, S.; Maguire, R. Pasture Biomass Estimation Using Ultra-High-Resolution RGB UAVs Images and Deep Learning. *Remote Sens.* **2023**, *15*, 5714. [CrossRef]
23. Johnson, I.R.; Chapman, D.F.; Snow, V.O.; Eckard, R.J.; Parsons, A.J.; Lambert, M.G.; Cullen, B.R. DairyMod and EcoMod: Biophysical pasture-simulation models for Australia and New Zealand. *Aust. J. Exp. Agric.* **2008**, *48*, 621–631. [CrossRef]
24. Vaze, J.; Johnston, W.H.; Teng, J.; Tuteja, N.K.; Johnson, I. Development and implementation of a generic pasture growth model (CLASS PGM). *Environ. Model. Softw.* **2009**, *24*, 107–114. [CrossRef]
25. Johnson, I.R.; Lodge, G.M.; White, R.E. The Sustainable Grazing Systems Pasture Model: Description, philosophy and application to the SGS National Experiment. *Aust. J. Exp. Agric.* **2003**, *43*, 711–728. [CrossRef]
26. Moore, A.D.; Holzworth, D.P.; Herrmann, N.I.; Huth, N.I.; Robertson, M.J. The Common Modelling Protocol: A hierarchical framework for simulation of agricultural and environmental systems. *Agric. Syst.* **2007**, *95*, 37–48. [CrossRef]
27. Keating, B.A. APSIM's origins and the forces shaping its first 30 years of evolution: A review and reflections. *Agron. Sustain. Dev.* **2024**, *44*, 24. [CrossRef]

28. Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S.; et al. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric. Syst.* **2017**, *155*, 269–288. [CrossRef] [PubMed]
29. Huang, J.; Tian, L.; Liang, S.; Ma, H.; Becker-Reshef, I.; Huang, Y.; Su, W.; Zhang, X.; Zhu, D.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [CrossRef]
30. Ogungbuyi, M.G.; Guerschman, J.; Fischer, A.M.; Crabbe, R.A.; Ara, I.; Mohammed, C.; Scarth, P.; Tickle, P.; Whitehead, J.; Harrison, M.T. Improvement of pasture biomass modelling using high-resolution satellite imagery and machine learning. *J. Environ. Manag.* **2024**, *356*, 120564. [CrossRef]
31. Liu, H.; Guo, P.; Liu, J.; Liu, R.; Tong, T. An Extension of Multiquadric Method Based on Trend Analysis for Surface Construction. IEEE J. *Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3435–3441. [CrossRef]
32. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
33. Nuss, W.A.; Titley, D.W. Use of Multiquadric Interpolation for Meteorological Objective Analysis. *Mon. Weather Rev.* **1994**, *122*, 1611–1631. [CrossRef]
34. Powell, M. Radial Basis Function Methods for Interpolation to Functions of Many Variables. *HERMIS Int. J. Comput. Maths Appl.* **2002**, 3.
35. Anjyo, K.; Lewis, J.P. RBF Interpolation and Gaussian Process Regression Through an RKHS Formulation. *J. Math. Ind.* **2011**, *3*, 63–71.
36. Mishra, P.K.; Nath, S.K.; Sen, M.K.; Fasshauer, G.E. Hybrid Gaussian-cubic radial basis functions for scattered data interpolation. *Comput. Geosci.* **2018**, *22*, 1203–1218. [CrossRef]
37. Jasek, K.; Pasternak, M.; Miluski, W.; Bugaj, J.; Grabka, M. Application of Gaussian Radial Basis Functions for Fast Spatial Imaging of Ground Penetration Radar Data Obtained on an Irregular Grid. *Electronics* **2021**, *10*, 2965. [CrossRef]
38. Smith, W.H.F.; Wessel, P. Gridding with continuous curvature splines in tension. *Geophysics* **1990**, *55*, 293–305. [CrossRef]
39. Török, P.; Lindborg, R.; Eldridge, D.; Pakeman, R. Grazing effects on vegetation: Biodiversity, management, and restoration. *Appl. Veg. Sci.* **2024**, *27*, e12794. [CrossRef]
40. Hassan, N.; Wang, Z. Paralleled grazing and mowing differentially affected plant community diversity and productivity in a semi-arid grassland. *Ecol. Process.* **2024**, *13*, 62. [CrossRef]
41. Cândido, B.; Mindala, U.; Ebrahimy, H.; Zhang, Z.; Kallenbach, R. Integrating Proximal and Remote Sensing with Machine Learning for Pasture Biomass Estimation. *Sensors* **2025**, *25*, 1987. [CrossRef]
42. Mutanga, O.; Masenyama, A.; Sibanda, M. Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. ISPRS J. Photogramm. *Remote Sens.* **2023**, *198*, 297–309. [CrossRef]
43. Jennewein, J.S.; Davis, B.W.; Seehaver-Eagen, S.; Nicolette, J.; Pittman, J.; Hively, W.D.; Goldsmith, A.; Hidalgo, C.; Reberg-Horton, C.; Mirsky, S.B. Multi-sensor proximal remote sensing for cover crop biomass estimation at high and moderate spatial resolutions. *Smart Agric. Technol.* **2025**, *12*, 101201. [CrossRef]
44. Filho, M.G.; Kuplich, T.M.; Quadros, F.L.F.D. Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *Int. J. Remote Sens.* **2020**, *41*, 2861–2876. [CrossRef]
45. Ogungbuyi, M.G.; Guerschman, J.; Fischer, A.M.; Mohammed, C.; Crabbe, R.A.; Harrison, M.T. Using vegetation indices from nanosatellites for timely prediction of pasture biomass. *Total Environ. Adv.* **2025**, *15*, 200130. [CrossRef]
46. Chen, Y.; Guerschman, J.; Shendryk, Y.; Henry, D.; Harrison, M.T. Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. *Remote Sens.* **2021**, *13*, 603. [CrossRef]
47. Nagler, P.L.; Inoue, Y.; Glenn, E.P.; Russ, A.L.; Daughtry, C.S.T. Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. *Remote Sens. Environ.* **2003**, *87*, 310–325. [CrossRef]
48. Cai, T.; Chang, C.; Zhao, Y.; Wang, X.; Yang, J.; Dou, P.; Otgonbayar, M.; Zhang, G.; Zeng, Y.; Wang, J. Within-season estimates of 10 m aboveground biomass based on Landsat, Sentinel-2 and PlanetScope data. *Sci. Data* **2024**, *11*, 1276. [CrossRef] [PubMed]
49. Tian, Z.; Fan, J.; Yu, T.; Leon, N.; Kaeppler, S.; Zhang, Z. Mitigating NDVI saturation in imagery of dense and healthy vegetation. ISPRS J. Photogramm. *Remote Sens.* **2025**, *227*, 234–250. [CrossRef]
50. Ranjbar, S.; Losos, D.; Dechant, B.; Hoffman, S.; Başakın, E.E.; Stoy, P.C. Harnessing Information From Shortwave Infrared Reflectance Bands to Enhance Satellite-Based Estimates of Gross Primary Productivity. *J. Geophys. Res. Biogeosci.* **2024**, *129*, e2024JG008240. [CrossRef]
51. Moisés, A.G.; Pascual, I.V.; González, J.J.I.; Zamarreño, C.R. Data Augmentation Techniques for Machine Learning Applied to Optical Spectroscopy Datasets in Agrifood Applications: A Comprehensive Review. *Sensors* **2023**, *23*, 8562. [CrossRef]
52. Ferber, F.F.; Gay, D.; Soulié, J.-C.; Diatta, J.; Maillard, O.-A. Kriging and Gaussian Process Interpolation for Georeferenced Data Augmentation. *arXiv* **2025**, arXiv:2501.07183. [CrossRef]

53. Kumar, A.; Sharma, A.; Singh, A.K.; Singh, S.K.; Saxena, S. Data Augmentation for Medical Image Classification Based on Gaussian Laplacian Pyramid Blending With a Similarity Measure. *IEEE J. Biomed. Health Inform.* **2023**, *29*, 3886–3893. [CrossRef] [PubMed]
54. Ortmann, M.; Buhmann, M. High accuracy quasi-interpolation using a new class of generalized multiquadrics. *J. Math. Anal. Appl.* **2024**, *538*, 128359. [CrossRef]
55. Sommariva, A.; Vianello, M. Random sampling and polynomial-free interpolation by Generalized MultiQuadrics. *J. Approx. Theory* **2025**, *306*, 106119. [CrossRef]
56. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [CrossRef]
57. Smith, H.D.; Dubeux, J.C.B.; Zare, A.; Wilson, C.H. Assessing Transferability of Remote Sensing Pasture Estimates Using Multiple Machine Learning Algorithms and Evaluation Structures. *Remote Sens.* **2023**, *15*, 2940. [CrossRef]
58. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.-F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [CrossRef]
59. Mutanga, O.; Skidmore, A. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. [CrossRef]
60. Mohammed, G.H.; Colombo, R.; Middleton, E.M.; Rascher, U.; van der Tol, C.; Nedbal, L.; Goulas, Y.; Pérez-Priego, O.; Damm, A.; Meroni, M.; et al. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sens. Environ.* **2019**, *231*, 111177. [CrossRef]
61. Zhang, Y.; Migliavacca, M.; Penuelas, J.; Ju, W. Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sens. Environ.* **2021**, *252*, 112121. [CrossRef]
62. Holzworth, D.; Huth, N.I.; Devoil, P.G.; Zurcher, E.J.; Herrmann, N.I.; McLean, G.; Chenu, K.; van Oosterom, E.J.; Snow, V.; Murphy, C.; et al. APSIM—Evolution towards a new generation of agricultural systems simulation. *Environ. Model. Softw.* **2014**, *62*, 327–350. [CrossRef]
63. Kyere, I.; Astor, T.; Graß, R.; Wachendorf, M. Multi-Temporal Agricultural Land-Cover Mapping Using Single-Year and Multi-Year Models Based on Landsat Imagery and IACS Data. *Agronomy* **2019**, *9*, 309. [CrossRef]

CHAPTER 6

Machine learning for grazing event detection and pasture utilisation quantification from sentinel-2 data

Estimating pasture availability solves only half the equation. Quantifying when and how much cows actually consume is essential for effective feed budgeting and pasture utilisation quantification, whilst knowing the precise timing of a grazing event matters as much as the biomass measurements themselves, since it is required to calculate rotation lengths, plan subsequent paddock allocations, determine rest periods for pasture recovery, and verify whether residual targets were met at the right point in the grazing cycle. However, temporal gaps from infrequent field measurements complicate accurate quantification of both. We address this challenge by utilising the interpolated pasture biomass data from Chapter 5. As cows graze a strip or paddock, pasture biomass declines in detectable patterns. In this chapter, we capture these patterns by integrating machine learning algorithms with anomaly detection methods to automatically identify grazing events from biomass trends. This framework delivers essential metrics for real-time pasture utilisation quantification across commercial dairy farms, enabling sustainable feed budgeting and informed grazing management decisions. The complete notebook and code for grazing detection implementation is available in the repository (chapter-6-grazing-detection/Chapter_6_Grazing_Detection_Notebook.ipynb). Key code excerpts are presented in Appendix D.

Smart Agriculture Technology (2026) 14, 101954

PUBLISHED MANUSCRIPT

The published version of this manuscript is included on the following page.

Smart Agricultural Technology 14 (2026) 101954

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Machine learning for grazing event detection and pasture utilisation quantification from sentinel-2 data

Blessing Nnenna Azubuike [a,d,*], Anna Chlingaryan [b,d], Martin Correa-Luna [a,d], Cameron E.F. Clark [c,d], Sergio C. Garcia [a,d]

[a] Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW 2570, Australia
[b] Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Camden, NSW 2570, Australia
[c] Gulbali Institute, Charles Sturt University, Wagga, NSW 2650, Australia
[d] Dairy UP Program, Camden, NSW 2570, Australia

ARTICLE INFO

Keywords:
Anomaly detection
Biomass
Cross-year validation
Semi-supervised learning
Temporal transferability

ABSTRACT

Effective pasture management in pasture-based dairy systems (P-BDS) requires accurate knowledge of when and where grazing occurs and how much biomass is consumed, yet manual recording methods are time consuming and prone to error. This study compared machine learning approaches for automated grazing event detection and pasture utilisation quantification from both rising plate meter and interpolated Sentinel-2 satellite data. Performance was evaluated against GPS-tracked cow grazing records across 12 commercial dairy farms in New South Wales, Australia (July 2022–June 2024). Eleven approaches were evaluated across within-year and cross-year validation scenarios to assess temporal transferability. Random Forest achieved optimal within-year detection performance (F1 score = 0.878, precision = 0.938, recall = 0.825), whilst One-Class Support Vector Machine (OCSVM) demonstrated superior cross-year transferability (F1 score = 0.692), outperforming supervised models by 7.6% on independent Year 2 data despite supervised models experiencing average performance degradation of 24.2%. Farm-level detection variability ranged from F1 = 0.500 to 0.815, with site-specific factors exerting stronger influence than regional characteristics. Independent pasture utilisation agreement validation on 218 GPS-confirmed events demonstrated strong biomass quantification concordance between satellite and ground measurements (pre-grazing biomass $R^2 = 0.966$; post-grazing biomass $R^2 = 0.998$; biomass removal $R^2 = 0.922$), establishing proof-of-concept that satellite systems can both identify when grazing occurs and accurately quantify biomass changes. Temporal alignment constraints between satellite revisit schedules and ground measurements limited validation to 19.4% of total events, highlighting the need for daily or near-daily observations through multi-sensor fusion approaches to achieve deployment in commercial P-BDS.

1. Introduction

Effective pasture management in dairy systems requires accurate knowledge of when and where grazing occurs, as this information directly influences feed budgeting, rotational grazing schedules and overall farm productivity [1,2]. In pasture-based dairy systems (P-BDS), the ability to detect grazing events (GE) rapidly and accurately, combined with precise quantification of pasture utilisation (PU), enables farmers to optimise forage management and improve economic returns. Pasture utilisation refers to the amount of herbage mass consumed during a GE, typically quantified as biomass removal (kg DM ha^{-1}) and harvesting rate (%), calculated as (pre-grazing biomass − post-grazing biomass) / pre-grazing biomass × 100. Accurate PU quantification enables farmers to optimise stocking rates, ensure adequate residuals for pasture recovery, balance feed demand with supply, and improve feed budgeting accuracy. Regular monitoring of herbage mass combined with accurate records of grazing timing and intensity can improve farm profitability by up to 15 % through better alignment of pasture availability with animal requirements and reduced wastage from overgrazing or under utilisation [1]. Traditional approaches to recording GE rely on manual documentation by farmers, a process that is time consuming, prone to error and often incomplete, particularly on larger operations managing multiple paddocks simultaneously [3]. As precision agriculture technologies advance, automated methods for detecting GE could

* Corresponding author.
E-mail address: blessing.azubuike@sydney.edu.au (B.N. Azubuike).

https://doi.org/10.1016/j.atech.2026.101954
Received 30 December 2025; Received in revised form 5 March 2026; Accepted 6 March 2026
Available online 7 March 2026

substantially reduce labour requirements whilst improving the accuracy and timeliness of PU records.

Wearable sensor technologies have emerged as potential solutions for livestock monitoring, with GPS tracking collars and accelerometers enabling detailed tracking of animal location and behaviour [4–6]. Accelerometers mounted on collars or ear tags capable of classifying cattle activities such as grazing, ruminating, resting and walking achieve classification accuracies exceeding 85 to 96 % for individual animal behaviours [7–10], but converting these measurements into paddock level GE detection requires additional processing, assumptions about herd dynamics and operational complexity for device management [11, 12].

Rising plate meters (RPM) are widely regarded as reliable tools for direct pasture biomass (PB) estimation because they convert compressed sward height readings into accurate measures of herbage mass [13,14]. However, they require manual operation, are labour intensive and lack scalability for large dairy enterprises, which limits both their temporal density and practicality for continuous monitoring [15–18]. These constraints have accelerated interest in remote sensing technologies, particularly satellite imaging platforms such as Sentinel 2, Planet and Landsat, which provide high spatial and temporal resolution suitable for monitoring PB at farm scale [16,19–21]. By leveraging freely available satellite data, farmers and researchers can establish continuous and detailed monitoring systems that support more informed pasture and grazing management decisions [22].

Satellite-based remote sensing (RS) provides an alternative approach for tracking vegetation dynamics without requiring animal-mounted sensors [23,24]. Satellite imagery excels at estimating standing PB through vegetation indices and spectral reflectance, with recent studies achieving accuracies of coefficient of determination (R^2) = 0.60 to 0.86 using machine learning (ML) approaches combined with Sentinel-2 or Landsat data [16,19,22,25]. The transition from PB estimation to GE detection represents a critical but underexplored application, as discrete livestock consumption events manifest as rapid PB disappearance patterns distinguishable from gradual senescence [26,27]. Recent interpolation techniques enable bridging temporal gaps in satellite observations, creating denser time series that better capture pasture growth dynamics [22,28]. Piecewise cubic Hermite interpolating polynomial and multiquadric (MQ) radial basis function (RBF) methods offer shape-preserving interpolation suitable for agricultural time series, maintaining biological plausibility whilst avoiding overshooting between measured points [29,30].

Machine learning approaches offer promising frameworks for classifying GE based on temporal PB patterns derived from satellite observations [15,25]. Whilst threshold based methods provide transparent solutions, ML algorithms often achieve superior performance by capturing complex nonlinear relationships [31–33]. Despite growing interest in automated pasture monitoring, significant gaps persist. Whilst ML applications in dairy farming have increased sevenfold since 2018, most focus on health monitoring rather than GE detection [33, 34]. Successful ML applications for detecting discrete livestock events have primarily focused on sheep parturition, whereas applications targeting dairy cattle grazing remain limited and understudied [35].

Critical knowledge gaps remain in four areas. First, comparative performance of different ML algorithms for GE detection remains unclear, with limited systematic evaluation of threshold-based, supervised learning and anomaly detection approaches applied to both satellite and ground-based measurement systems. Second, temporal transferability of trained classifiers across seasons and production years needs assessment for operational deployment particularly when accounting for inherent timing uncertainties between measurement systems. Third, optimal validation frameworks for automated detection systems require development, specifically addressing how to appropriately match predictions against GPS ground truth when measurement timing varies between data sources. Fourth, the agreement between satellite-derived and traditional ground-based PU measurements requires quantification to determine whether remote sensing can reliably substitute for labour-intensive manual monitoring of pre-grazing PB, post-grazing residuals and removal amounts across diverse farming conditions. This research tests the hypothesis that the rate of PB disappearance between consecutive satellite observations, derived from interpolated time series, can simultaneously detect GE and accurately quantify PU in commercial dairy systems without requiring manual farm records.

This research addresses these gaps through two complementary objectives. The first objective is to develop and systematically compare ML-based approaches for automated GE detection from both RPM and interpolated Sentinel-2 satellite data, evaluating their performance against GPS-recorded grazing timing (ground truth or 'gold standard') across multiple production years to assess temporal transferability and operational robustness. The second objective is to independently quantify the agreement between RPM-derived and satellite-derived PU measurements (pre-grazing PB, post-grazing residuals, and removal amounts) for GPS-confirmed GE, assessing whether satellite systems can accurately estimate PB changes when grazing actually occurs. By separating detection performance from measurement agreement, this framework enables independent evaluation of whether satellite systems can reliably (1) identify when grazing occurs and (2) accurately quantify PB changes, providing empirical evidence for their deployment capabilities in commercial P-BDS.

2. Materials and methods

2.1. Study sites and data sources

This study utilised data from 12 commercial dairy farms across New South Wales, Australia, collected between July 2022 and June 2024. The farms were distributed across three coastal regions: mid-coast (3 farms, latitude range: -34.03° to -31.72° S, longitude range: 150.65° to 152.68° E), north coast (4 farms, latitude range: -28.90° to -28.68° S, longitude range: 152.91° to 153.13° E), and south coast (5 farms, latitude range: -36.82° to -36.64° S, longitude range: 149.60° to 149.90° E). Regional climatic conditions varied substantially, with long term mean annual rainfall averaging approximately 780 mm in the south coast region, 1284 mm in the mid coast district and 1073 mm in the north coast area. Daily air temperatures ranged from approximately 5°C to 20°C during winter months and 20°C to 35°C throughout summer periods. The study monitored 60 paddocks across the twelve farms, with five paddocks per farm designated for intensive ground truth monitoring throughout the study period. Farm scale utilisable grazing areas ranged from 64.5 to 313 ha. All participating farms operated primarily with kikuyu (*Cenchrus clandestinus)* pasture oversown annually with annual ryegrass (*Lolium multiflorum* L.) during the autumn and winter months. Fig. 1 shows the geographic distribution of the 12 farms across the three coastal regions used in this study.

2.1.1. Grazing dataset from GPS tracking collars

GPS collars were fitted to three lactating dairy cows per farm throughout the study period to provide ground truth GE records. GPS trackers (Yabby Edge Cellular, Digital Matter, Australia) utilising IoT cellular network bands (LTE-M CatM1 and NB-IoT) recorded location coordinates at 60-min intervals with positional accuracy of ±2.5 m under optimal satellite visibility [36]. GPS data underwent quality control procedures to exclude invalid coordinates (latitude/longitude outside farm boundaries), stationary positioning errors and satellite signal dropouts. Paddock occupancy status was determined by spatial intersection of GPS coordinates with farm paddock boundaries using geographic information system operations. A paddock was classified as grazed when at least one collared animal remained within paddock boundaries for a minimum continuous duration of 2 h, consistent with typical dairy grazing rotation periods. GE start dates were defined as the first day when paddock occupancy criteria were met, and end dates as the last day of continuous occupancy before herd movement to a

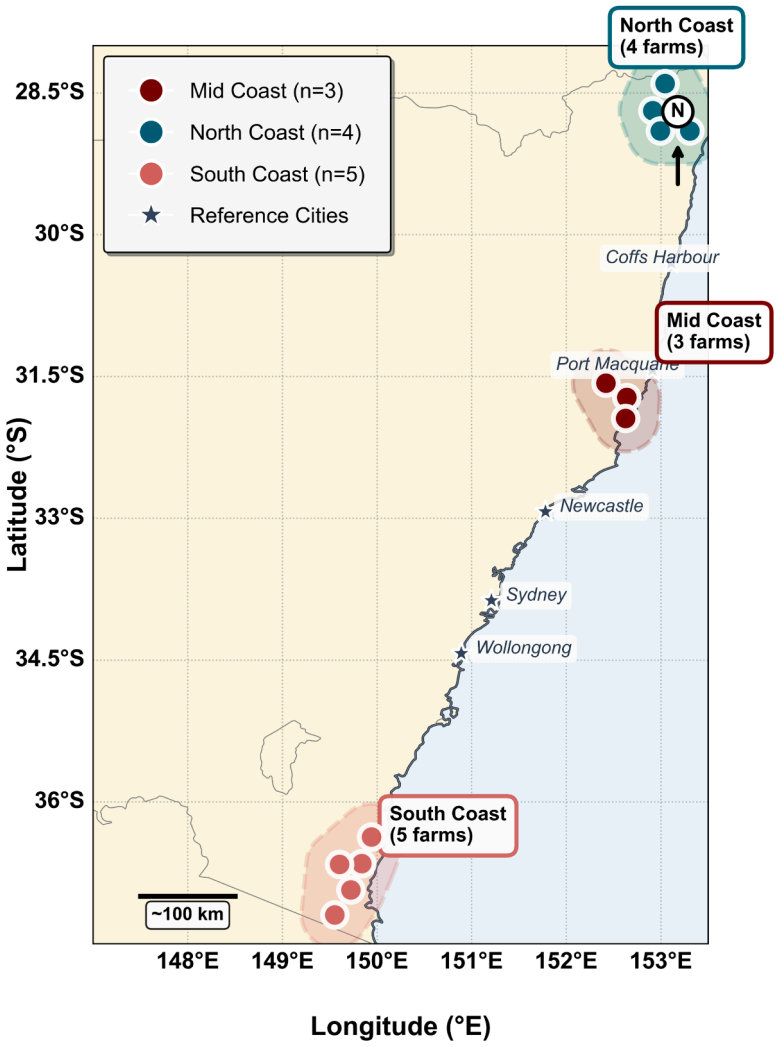


Fig. 1. Distribution of 12 commercial dairy farms across three coastal regions of New South Wales, Australia: Mid Coast (n=3), North Coast (n=4), South Coast (n=5). Reference cities shown for geographic context.

different paddock. Consecutive days when GPS records indicated cattle presence in the same paddock were grouped into single GE to prevent over-counting. Dates separated by ≤5 days were consolidated into unified events, with event start defined as the earliest date and event end as the latest date within each group. This grouping procedure generated 1492 GEs across the 60 monitored paddocks throughout the 2-year study period (July 2022 to June 2024), with mean event duration of 2.6 days (median 2.0 days, range 1–55 days). These GPS-derived events provided independent ground truth for model training and validation.

The use of three GPS collars per farm was based on the manufacturer recommendations for herd-level tracking in intensive dairy systems [36]. In the pasture-based dairy industry of Australia, cows are managed as cohesive herds or mobs that are brought from paddocks to the dairy for milking twice daily and returned to designated grazing allocations after each milking. These allocations are typically small areas with high grazing intensity (commonly approximately 100 cows/ha as instantaneous stocking rate), ensuring that animals remain in close proximity throughout grazing periods. Under these management conditions, a single GPS-collared animal would theoretically provide adequate herd-level location data. However, three collars per farm were deployed to account for potential device failures, signal dropouts, or temporary separation of individual animals during milking or health treatments, thereby ensuring continuous paddock occupancy monitoring throughout the study period.

2.1.2. *Rising plate meter measurements*

Ground truth PB measurements were conducted using a Jenquip EC20 Electronic Pasture Meter (Feilding, New Zealand). Monthly calibration activities were implemented on two designated paddocks per farm [37], involving collection of nine 0.1 m^2 quadrat samples cut to 5 cm stubble height. Samples were dried for 48 h at 65°C before weighing. Linear regression analysis between compressed sward height (CSH) readings and measured dry matter values provided farm specific calibration equations for converting height measurements to PB estimates expressed as kg DM ha^{-1}. For each monitoring event, a minimum of 70 individual measurements were recorded within each paddock using systematic sampling procedures. Measurements were averaged to provide a single PB estimate per paddock per monitoring date. Quality control procedures excluded values outside the 1000 to 4000 kg DM ha^{-1} range appropriate for the study pasture systems. RPM measurements were collected weekly during Year 1 (July 2022 to June 2023) and fortnightly during Year 2 (July 2023 to June 2024) [37]. Following filtering to the 60 common paddocks monitored across both years, the RPM dataset comprised 3066 observations (Year 1: 2135 observations; Year 2: 931 observations). After removing 120 observations with incomplete temporal features, 2946 observations remained for model development and validation. These RPM measurements were needed to provide independent ground truth data for validating satellite-based GE detection timing, and enabling quantification of PU metrics including

pre-grazing PB, post-grazing PB and removal amounts.

2.1.3. Sentinel-2 satellite data

Sentinel-2 imagery was acquired throughout the study period, providing surface reflectance data at 10 m spatial resolution with a nominal 5-day revisit frequency. Nine spectral bands were utilised: blue (band 2, 490 nm), green (band 3, 560 nm), red (band 4, 665 nm), near infrared (band 8, 842 nm), three red edge bands (bands 5, 6, 7 at 705, 740 and 783 nm respectively), and two shortwave infrared bands (bands 11 and 12 at 1610 and 2190 nm respectively). Imagery pre-processing included atmospheric correction using the Sen2Cor processor, resampling to a consistent 10 m grid and geometric correction. Quality assurance utilised the Function of mask (Fmask) algorithm to exclude pixels affected by clouds, shadows, water or snow [38]. For each paddock and acquisition date, spectral statistics were calculated by averaging valid pixel values within paddock boundaries after removing outliers falling outside 1.5 interquartile ranges from the median. Paddock-level Normalised Difference Vegetation Index (NDVI) values were averaged across all valid pixels within each paddock boundary for each satellite acquisition date.

Pasture biomass estimates derived from satellite observations were obtained using a previously developed monitoring framework [22]. That framework addressed temporal gaps in RPM measurements using multiquadric (MQ) radial basis function interpolation to create gap-free training data. An XGBoost machine learning model was then trained to predict PB using Sentinel-2 spectral bands, derived vegetation indices (NDVI, EVI, SAVI, NDRE), weather variables (temperature, rainfall, solar radiation, evapotranspiration), and seasonal characteristics, achieving cross-validated performance of $R^2 = 0.70$ and RMSE = 293 kg DM ha^{-1}. This approach generated PB estimates by combining direct RPM measurements, satellite observations, and interpolated values. The current study utilised this interpolated satellite-derived dataset from 12 of the 16 original farms in [22]. Farms were included if they had: (1) continuous GPS collar deployment throughout both study years, (2) minimum 5 observations per paddock with valid PB data (1000–4000 kg DM ha^{-1}), and (3) sufficient temporal overlap between PB measurements and GPS-confirmed GE for model validation. Four farms failed to meet these criteria. Following filtering to the 60 common paddocks monitored throughout both study years, the dataset comprised 7511 observations across the study period (Year 1: 4264 observations; Year 2: 3247 observations). After removing 120 observations with incomplete temporal features, 7391 observations remained for model development and validation. All interpolated records were tagged to enable downstream distinction from directly observed values. The MQ interpolation was applied to the full paddock biomass time series prior to temporal splitting into Year 1 and Year 2 subsets. Whilst the constrained nature of Multiquadric interpolation minimises cross-boundary information leakage, future implementations should apply interpolation independently within each temporal subset.

2.2. Data preprocessing and feature engineering

Two primary datasets were constructed for model development, one containing only direct RPM measurements (RPM dataset), and one containing interpolated PB estimates from Sentinel-2 satellite data (MQ dataset). The RPM dataset comprised 2946 observations across 60 paddocks (Year 1: 2075 observations; Year 2: 871 observations), whilst the MQ dataset comprised 7391 observations across the same paddocks (Year 1: 4204 observations; Year 2: 3187 observations). Each dataset retained only observations with valid PB estimates between 1000 and 4000 kg DM ha^{-1}. A critical methodological challenge in automated GE detection involves accounting for inherent temporal misalignment between measurement systems. GPS collars record exact paddock entry dates, whereas RPM measurements occur at weekly to fortnightly intervals and satellite observations depend on cloud-free acquisition schedules. To address this temporal uncertainty, a tolerance-based labelling approach was implemented where for each observation in the RPM and MQ datasets, the temporal proximity to GPS-recorded GE was calculated. An observation was labelled as *grazed* if a GPS-confirmed GE occurred within ±7 days of the observation date on the same paddock. This 7-day tolerance window accounts for typical measurement frequencies in operational dairy systems whilst maintaining biological relevance to the GE. Observations falling outside this window for all GPS events were labelled *not grazed* to create a realistic validation framework that acknowledges measurement timing uncertainties.

For each observation at time t within each paddock, temporal features were constructed using forward-looking windows. The next observation date, next PB value, and time interval (days) between consecutive observations were calculated using grouped temporal shifts within each paddock time series. Biomass change (ΔBiomass) was computed as the difference between consecutive measurements, and the PB disappearance rate was calculated as:

$$\text{Disappearance Rate} = \frac{\max(0, -\Delta\text{Biomass})}{\Delta t}$$

where $\Delta Biomass$ represents the change in PB (kg DM ha^{-1}) between consecutive observations as a result of a GE, Δt is the time interval in days. Negative changes (biomass loss) were clipped to zero to focus specifically on disappearance events, as PB increases indicate growth rather than consumption. This rate metric normalises biomass loss by the inter-observation interval, enabling fair comparison between observations with varying temporal spacing.

Additional temporal and biomass-related features were engineered to capture growth patterns and grazing signals. Temporal features comprised day of year (1–366) to capture phenological patterns, week of year (1–52) for seasonal grazing intensity, month (1–12) as a categorical variable for broader seasonal effects, and year (2022–2024) for inter-annual climatic variability. Categorical features included farm identifier, paddock code, region (Mid Coast, North Coast, South Coast) and season (spring, summer, autumn, winter) to capture spatial and climatic variation across the study area.

2.3. Model development, training, and evaluation

Model performance was evaluated using two temporal validation scenarios designed to assess robustness and transferability. The within-year scenario used a temporal split on 1 January 2023, with all observations before this date (July 2022 to December 2022) used for training, and all observations after this date (January 2023 to June 2024) used for independent validation. This scenario assessed model generalisation across seasons and production years within the temporal span of available data. The cross-year scenario trained models exclusively on Year 1 data (July 2022 to June 2023) and tested on independent Year 2 data (July 2023 to June 2024), representing the most stringent test of temporal transferability across annual production cycles. Both scenarios maintained strict temporal independence between training and testing data, with validation periods representing previously unseen temporal conditions. Four validation datasets were constructed: RPM within-year, RPM cross-year, MQ within-year, and MQ cross-year. Each dataset was independently split into training and testing periods according to the scenario temporal boundaries. This design enabled systematic comparison of detection performance between measurement systems (RPM vs MQ) and temporal contexts (within-year vs cross-year).

2.4. Threshold-based classification

The threshold-based approach employed a single continuous predictor (PB disappearance rate) to discriminate between grazed and not-grazed intervals. For a given threshold τ, an observation was classified as grazed if its disappearance rate exceeded τ, otherwise classified as not grazed. The optimal threshold was determined independently for each

validation dataset by minimising the total absolute error (TAE) between predicted and actual GE counts aggregated at the paddock level. For each paddock p , let A_p represent the actual number of GE and $P_p\ (\tau)$ the predicted number of GE at threshold τ. The TAE was computed as:

$$\mathrm{TAE}(\tau) = \sum_{p=1}^{N} \left| A_p - P_p(\tau) \right|$$

where N is the total number of paddocks. This objective function was minimised by evaluating 200 equally spaced candidate thresholds spanning the full range of observed disappearance rates. The threshold yielding the minimum TAE was retained as optimal and used for all subsequent predictions on that dataset.

2.5. Supervised machine learning and anomaly detection approaches

Nine supervised ML algorithms were systematically evaluated for GE classification: Random Forest, Extra Trees, XGBoost, Gradient Boosting, Logistic Regression, Classification and Regression Trees (CART), K-Nearest Neighbours, Naive Bayes, and Support Vector Machine (SVM). All continuous features were standardised using z-score normalisation prior to model training to ensure comparable feature scales across different measurement units. Categorical features were one-hot encoded for compatibility with tree-based and linear algorithms. Hyperparameter optimisation was conducted using 3-fold cross-validation on the training data with F1 score as the optimisation metric. The F1 score is the harmonic mean of precision (proportion of predicted GE that were correct) and recall (proportion of actual GE successfully detected), calculated as:

$$F_1 = \frac{2 \times \mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

F1 ranges from 0 (worst) to 1 (best) and provides a balanced measure of classification performance that accounts for both false positives and false negatives.

Grid search explored conservative hyperparameter ranges to balance model complexity with generalisation performance. For tree-based methods (Random Forest, Extra Trees, Gradient Boosting), the number of estimators ranged from 100 to 300, maximum tree depth from 5 to 15, minimum samples per split from 2 to 10, and minimum samples per leaf from 1 to 4. For XGBoost, learning rates of 0.01, 0.05 and 0.1 were evaluated alongside subsample ratios of 0.8 and 1.0. SVM explored regularisation parameters (C) from 0.1 to 100 with radial basis function kernel and gamma values of 'scale' and 'auto'. K-Nearest Neighbours evaluated neighbour counts from 3 to 15 with both uniform and distance-weighted voting. For each algorithm, the hyperparameter configuration achieving the highest cross-validation F1 score was selected. The optimised model was then retrained on the full training dataset and evaluated on the independent temporal validation data.

Anomaly detection was implemented using One-Class Support Vector Machine (OCSVM) under a semi-supervised paradigm. The OCSVM was trained exclusively on observations labelled *not grazed* in the training data, treating GE as anomalies to be detected. This approach mirrors operational scenarios where GE are infrequent relative to normal pasture growth periods. The OCSVM learns a decision boundary around normal observations in a transformed feature space using a radial basis function kernel. Observations falling outside this boundary during testing are flagged as anomalies (potential GE). The contamination parameter, which controls the expected proportion of anomalies in the training data, was set to 0.05 to account for potential mislabelled observations whilst maintaining model sensitivity. The kernel coefficient (gamma) was set to 'scale' (1/(n_features × variance)), and the regularisation parameter (nu) was optimised via cross-validation within the range [0.01, 0.1, 0.2]. Predictions from the OCSVM were converted to binary classifications, with anomalies (output = -1) classified as GE and normal observations (output = +1) classified as not grazed. All supervised and semi-supervised ML models were implemented using the scikit-learn library [39] in Python (version 1.7.2).

2.6. Performance evaluation framework

2.6.1. Part A: grazing event detection performance

Detection performance was evaluated using both observation-level and event-level metrics. Observation-level metrics treated each measurement as an independent classification decision and included precision, recall, F1 score, and accuracy, calculated from confusion matrix elements where TP = true positives (correctly predicted grazed), TN = true negatives (correctly predicted not grazed), FP = false positives (incorrectly predicted grazed), and FN = false negatives (incorrectly predicted not grazed). Event-level detection assessment compared predicted GE against GPS ground truth within the ±7 day temporal tolerance window. For each GPS-confirmed event, a true positive detection was recorded if any observation within the tolerance window was classified as grazed. GPS events with no detections within the window were counted as false negatives. Predicted GE not matching any GPS event within the tolerance window were counted as false positives. This evaluation approach accounts for temporal uncertainty whilst rewarding models that detect GE within operationally relevant time windows. Confusion matrices were constructed for each validation scenario (RPM within-year, RPM cross-year, MQ within-year, MQ cross-year) showing the distribution of true positives, false positives, false negatives and true negatives. Farm-level detection performance was assessed by calculating F1 scores independently for each farm, enabling identification of regional and operational differences in model performance.

2.6.2. Part B: pasture utilisation agreement analysis

The second evaluation component assessed agreement between RPM and satellite-derived PU measurements for GPS-confirmed GEs, independently of detection model performance. This analysis addressed whether satellite systems can accurately quantify PB changes when grazing actually occurs, separating measurement agreement from detection accuracy. For each GPS-confirmed GE, pre-grazing and post-grazing PB measurements were extracted from both RPM and MQ datasets using biologically informed temporal windows. Pre-grazing cover was defined as the peak PB observation occurring within a search window bounded by: (1) the end date of the previous GE on the same paddock plus 1 day, or (2) 21 days before the current event start date if no previous event existed or if the inter-event gap exceeded 21 days. This window captures the maximum pasture accumulation before grazing whilst avoiding contamination from previous grazing periods. Post-grazing PB cover was defined as the minimum PB observation occurring between the GE end date and up to 5 days afterward, capturing the lowest residual PB immediately following livestock removal.

Events were excluded from PU analysis if either pre-grazing or post-grazing observations were unavailable within the specified windows, or if calculated removal (pre-grazing minus post-grazing biomass) was negative, indicating measurement errors or data quality issues. For events with valid measurements from both RPM and MQ datasets, the following PU metrics were calculated: **Pre-grazing PB** (kg DM ha^{-1}): Peak PB before grazing ; **Post-grazing PB** (kg DM ha^{-1}): Minimum residual after grazing; **Removal** (kg DM ha^{-1}): Pre-grazing minus post-grazing biomass.

Agreement between PB estimates using RPM and satellite-derived measurements was quantified using Pearson correlation coefficient (r), RMSE, and bias. Correlation assessed the strength of linear relationship between measurement systems. RMSE quantified absolute disagreement:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\text{RPM}_i - \text{Satellite}_i)^2}{n}}$$

Bias quantified systematic over- or under-estimation by satellite measurements:

$$\text{Bias} = \frac{\sum_{i=1}^{n}(\text{Satellite}_i - \text{RPM}_i)}{n}$$

where n is the number of matched events. Positive bias indicated satellite over-estimation relative to RPM, whilst negative bias indicated under-estimation. Agreement analysis was conducted separately for pre-grazing PB, post-grazing PB and removal to identify which PU components exhibited strongest concordance between measurement systems. Farm-level agreement was assessed by calculating correlation coefficients independently for farms with ≥3 matched events, enabling identification of site-specific factors influencing measurement concordance.

All analyses were conducted in Python (version 3.11.4). For each validation scenario, model performance metrics were computed independently on training and testing periods. Statistical significance of correlations was assessed using two-tailed *t*-tests with significance threshold $\alpha = 0.05$. Comparison between RPM and MQ datasets enabled quantitative assessment of operational value added by temporal densification through satellite interpolation relative to sparse field measurements alone.

3. Results

The detection validation framework comprised four datasets spanning two production years. Year-to-year comparison revealed increased grazing prevalence (51.3–51.4 % in Year 1 vs. 57.6–58.3 % in Year 2) and declining mean PB (2730–2747 vs. 2601–2616 kg DM ha^{-1}), consistent with variable seasonal conditions. Both RPM and MQ datasets captured similar biomass ranges (1135–3860 kg DM ha^{-1}), providing comparable feature distributions despite differing temporal resolutions.

3.1. Grazing event detection performance

3.1.1. Model performance comparison

Eleven detection approaches were evaluated across four validation scenarios (RPM/MQ × Within-Year/Cross-Year), yielding 56 model configurations. Random Forest achieved the highest within-year performance for both RPM (F1 = 0.878, precision = 0.938, recall = 0.825) and MQ datasets (F1 = 0.853, precision = 0.885, recall = 0.823). For cross-year validation, OCSVM demonstrated superior temporal transferability, achieving F1 = 0.692 for RPM and F1 = 0.626 for MQ (Table 1).

Performance degradation from within-year to cross-year validation averaged 24.2 % for supervised models (RPM: F1 decline from 0.747 to 0.562; MQ: F1 decline from 0.733 to 0.518) (Fig. 3). Tree-based ensemble methods (Random Forest, XGBoost, Gradient Boosting, Extra Trees) consistently outperformed other model families. Within-year scenarios showed 9 of the top 10 configurations, with F1 scores ranging from 0.752 to 0.878, all from tree-based models applied to either RPM or MQ data (Fig. 2). Threshold-based approaches achieved maximum F1 of 0.554 (High Sensitivity, RPM Cross-Year), whilst linear models (Logistic Regression) and probabilistic classifiers (Naive Bayes) demonstrated limited effectiveness (F1 < 0.622 across all scenarios).

OCSVM exhibited contrasting behaviour, performing poorly within-year (F1 = 0.321 for RPM, 0.291 for MQ) but substantially better cross-year (F1 = 0.692 for RPM, 0.626 for MQ), outperforming all supervised models in temporal transferability assessment. The comprehensive performance heatmap revealed distinct patterns across all 56 configurations, with tree-based methods showing consistent strong performance (F1 > 0.75) in within-year scenarios but substantial degradation in cross-year contexts (Fig. 4). RPM-based models marginally outperformed MQ-based models in cross-year scenarios (average F1 difference: +0.044), with 12 of 14 model types showing higher RPM performance. Random Forest, Extra Trees, and XGBoost maintained F1 > 0.638 on RPM cross-year validation, compared to F1 > 0.602 for the same models on MQ data.

Table 1
Model performance and independent variables.

Part A: Best Model Performance by Validation Scenario

Validation Scenario	Data Source	Model	Precision	Recall	F1	Accuracy
Within-Year	RPM	Random Forest	0.938	0.825	0.878	0.882
Within-Year	MQ	Random Forest	0.885	0.823	0.853	0.854
Cross-Year	RPM	OCSVM	0.61	0.799	0.692	0.584
Cross-Year	MQ	OCSVM	0.599	0.655	0.626	0.549

Part B: Independent Variables Used in Model Training

Feature Category	Feature Name	Description
Biomass (n=2)	Pasture Biomass (PB)	Current observation pasture biomass (kg DM/ha)
	Next PB	Next observation pasture biomass (kg DM/ha)
Temporal (n=9)	day_of_year	Day of year (1–365)
	week_of_year	Week of year (1–52)
	Week	Calendar week number
	Month	Calendar month (1–12)
	Year	Calendar year
	Season_Autumn	Binary indicator for autumn season
	Season_Spring	Binary indicator for spring season
	Season_Summer	Binary indicator for summer season
	Season_Winter	Binary indicator for winter season
Derived (n=5)	days_interval	Time interval between consecutive observations (days)
	biomass_change	Change in biomass between observations (kg DM/ha)
	biomass_change_abs	Absolute biomass change (kg DM/ha)
	disappearance_rate	Biomass change as percentage of initial biomass (%)
	rate_per_day	Daily rate of biomass change (kg DM/ha/day)

Note: Within-Year scenarios trained and tested on different temporal periods within Year 1 (Jul 2022 - Jun 2023). Cross-Year scenarios trained on Year 1 and tested on Year 2 (Jul 2023 - Jun 2024). RPM = Rising Plate Meter dataset; MQ = Multiquadric interpolated satellite dataset. All models used the same 13 independent variables (after one-hot encoding of Season into 4 binary indicators). No feature selection procedure was performed. Models were implemented using scikit-learn library in Python.

3.1.2. Temporal transferability assessment

Random Forest demonstrated strong within-year performance on the RPM dataset, correctly classifying 878 of 1064 grazed observations (true positives) and 953 of 1011 non-grazed observations (true negatives), whilst misclassifying 58 as false positives and 186 as false negatives (accuracy = 0.882). However, when this model was applied to cross-year validation, performance degraded substantially (F1 decline from 0.878 to 0.642), consistent with the 24.2 % average degradation observed across supervised models (Fig. 3). OCSVM exhibited contrasting temporal characteristics, achieving F1 = 0.692 on cross-year validation despite weaker within-year performance. The cross-year confusion matrix revealed 406 true positives and 103 true negatives against 260 false positives and 102 false negatives (accuracy = 0.584). Critically, OCSVM outperformed Random Forest by 0.049 F1 points (7.6

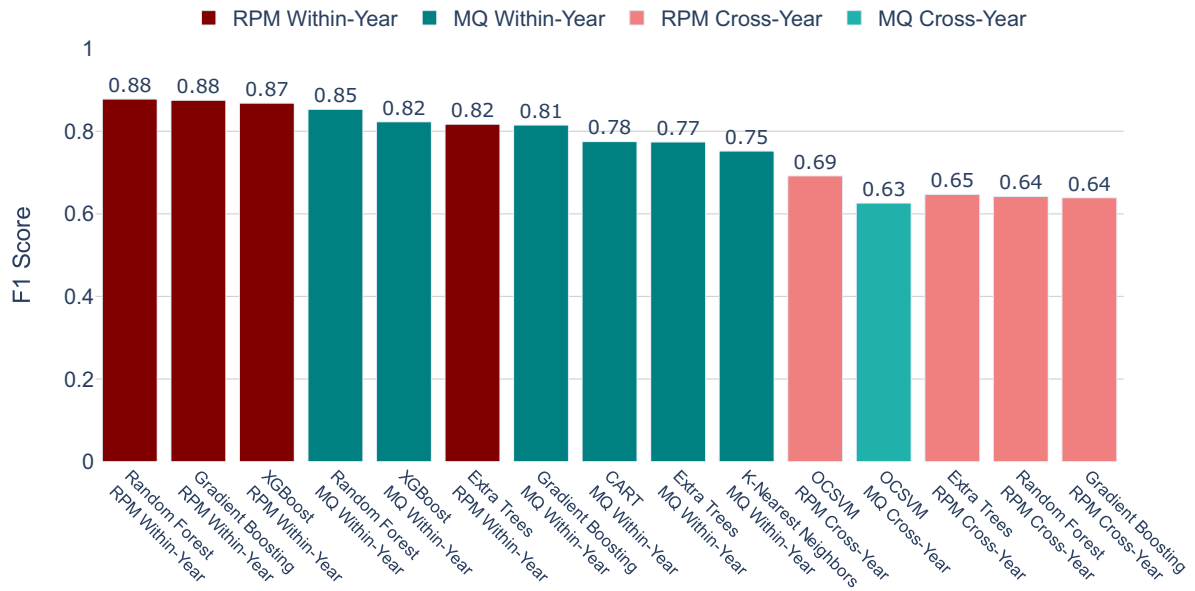


Fig. 2. Performance comparison of the top 15 model configurations for grazing event detection, ranked by F1 score. Bars are coloured by validation scenario: maroon (RPM within-year), teal (MQ within-year), light coral (RPM cross-year), and light seagreen (MQ cross-year). RPM = Rising Plate Meter; MQ = Multiquadric interpolated satellite estimates.

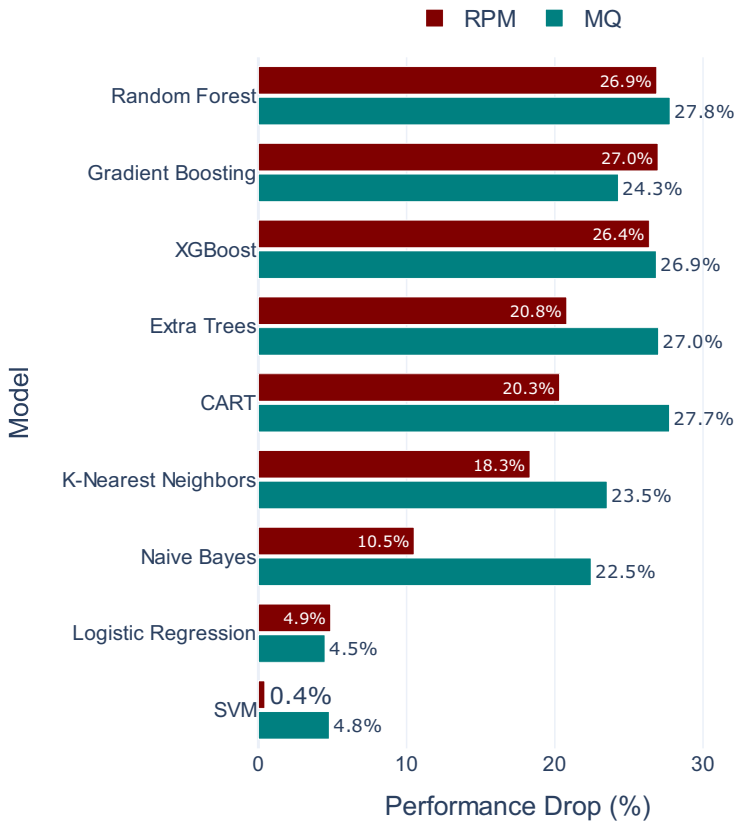


Fig. 3. Performance degradation from within-year to cross-year validation. Percentage decline in F1 score when models trained on Year 1 (Jul 2022–Jun 2023) were tested on Year 2 (Jul 2023–Jun 2024). Maroon bars: RPM dataset; teal bars: MQ = Multiquadric interpolated satellite estimates dataset. RPM = Rising Plate Meter; MQ = Multiquadric interpolated satellite estimates.

% relative improvement) when both models were evaluated on Year 2 test data, demonstrating superior generalisation to novel temporal conditions. This advantage stemmed from OCSVM higher recall (0.799 vs 0.661) at the cost of reduced precision (0.610 vs 0.625).

Cross-year evaluation across all 10 detection approaches revealed three distinct performance tiers. The anomaly detection method (OCSVM, F1 = 0.692) achieved the highest transferability. Tree-based ensemble methods formed a second tier (Extra Trees, Random Forest,

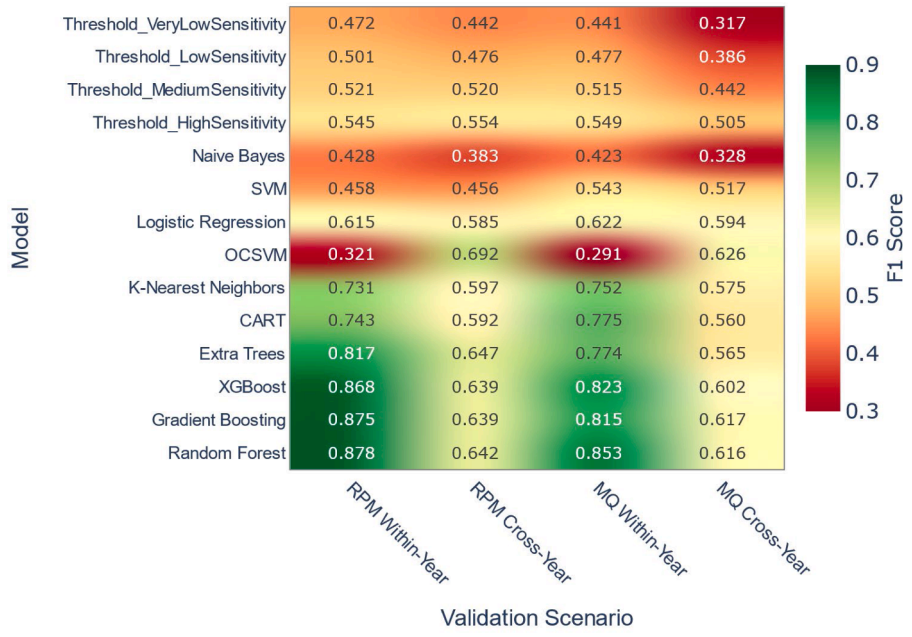


Fig. 4. Comprehensive performance heatmap showing F1 scores for all 56 model configurations (14 models × 4 validation scenarios). Cell colours range from red (poor, F1 < 0.4) to green (strong, F1 > 0.8). RPM = Rising Plate Meter; MQ = Multiquadric interpolated satellite estimates.

Gradient Boosting, XGBoost: F1 = 0.639–0.647), maintaining moderate but degraded performance. Distance-based, linear, and probabilistic classifiers demonstrated poor temporal transferability (K-Nearest Neighbours, CART, Logistic Regression, SVM, Naive Bayes: F1 = 0.383–0.597), with SVM and Naive Bayes particularly affected by temporal distribution shifts.

3.1.3. *Farm-level detection variability*

Farm-level performance exhibited substantial heterogeneity across the 12 farms, with F1 scores ranging from 0.500 to 0.815 (mean = 0.675, SD = 0.084, coefficient of variation = 0.124) when applying OCSVM to cross-year validation data (Fig. 5). The highest-performing farm (Farm 8, Mid Coast) achieved F1 = 0.815 across 65 observations (46 grazed), whilst the lowest-performing farm (Farm 4, North Coast) recorded F1 = 0.500 across 34 observations (27 grazed), representing a 63 % relative difference in detection effectiveness. Regional patterns emerged when aggregating farm-level performance. Mid Coast farms (n = 3: Farms 2, 8, 9) demonstrated the strongest detection performance (mean F1 = 0.735 ± 0.072, range 0.676–0.815) across 211 total observations with 58.8 % grazing prevalence. North Coast farms (n = 4: Farms 3, 4, 6, 12) exhibited intermediate performance (mean F1 = 0.662 ± 0.113, range 0.500–0.761) across 257 observations with 64.6 % grazing prevalence,

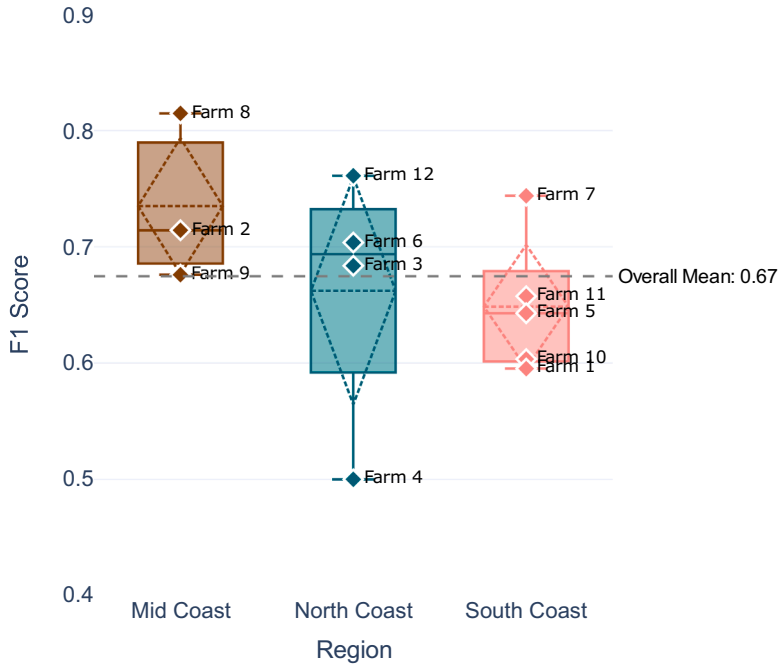


Fig. 5. Farm-level detection performance variability for OCSVM cross-year validation across the 12 commercial dairy farms grouped by coastal region. Boxplots show median, interquartile range, and range, with mean indicated by diamond symbol. Horizontal dashed line indicates overall mean (F1 = 0.675).

though with notably higher variability (SD = 0.113) than other regions. South Coast farms (n = 5: Farms 1, 5, 7, 10, 11) recorded the lowest regional mean (F1 = 0.649 ± 0.059, range 0.595–0.744) across 403 observations despite the most consistent performance (SD = 0.059) and intermediate grazing prevalence (54.1 %). The substantial between-farm variability (range = 0.315 F1 points) exceeded the between-region variability (range = 0.086 F1 points between regional means), indicating that farm-specific factors exerted stronger influence on detection performance than broad regional characteristics. Within the North Coast region, performance ranged from F1 = 0.500 (Farm 4) to F1 = 0.761 (Farm 12), demonstrating that high- and low-performing operations coexisted within the same climatic and geographic conditions.

3.1.3.1. Drivers of farm-level performance variability. To identify factors influencing detection performance heterogeneity, we examined correlations between farm-level F1 scores and five operational characteristics across the 12 study farms. Grazing intensity (mean stocking rate during lactating herd grazing: 82.2 cows/ha, range: 43.0–161.4 cows/ha) exhibited a moderate negative correlation with F1 scores (r = -0.504), suggesting farms operating at higher stocking densities experienced reduced detection accuracy. This relationship may reflect more rapid biomass depletion under intensive grazing, which compresses the temporal detection window available for satellite observation between grazing events.

Observation density (mean: 14.7 observations per paddock, range: 5.4–22.4) showed a moderate positive correlation with F1 scores (r = 0.392), indicating farms with denser temporal sampling achieved marginally better detection performance. However, paddock area (mean: 3.6 ha, range: 1.3–6.7 ha; r = -0.028), grazing prevalence (mean: 58.3 % of observations classified as grazed, range: 44.4–79.4 %; r = 0.188), and mean grazing duration (mean: 2.5 days, range: 1.5–4.7 days; r = 0.032) demonstrated negligible relationships with detection accuracy. These patterns suggest that management intensity (stocking rate) may influence detection performance more strongly than spatial characteristics, temporal sampling frequency, or individual event duration alone, though the correlational nature of this analysis and small sample size (n=12 farms) preclude definitive causal inference.

3.2. Pasture biomass and utilisation agreement

Agreement analysis included 218 of 1124 GPS validation events (19.4 %) where concurrent pre- and post-grazing measurements were available for both RPM and satellite systems. Exclusion occurred randomly due to measurement timing constraints (484 events lacked post-grazing RPM observations, 370 lacked pre-grazing observations), not based on event characteristics. The resulting sample spanned both production years (Year 1: 84 %, Year 2: 16 %) with seasonal representation across summer (39 %), autumn (39 %), spring (15 %), and winter (7 %).

Exclusion rates varied seasonally but remained systematically high across all periods. Autumn exhibited the highest inclusion rate (24.1 %, 85 of 352 events), followed by summer (21.2 %, 85 of 400), spring (15.5 %, 32 of 207), and winter (9.7 %, 16 of 165). Despite seasonal variation, all periods demonstrated exclusion rates exceeding 75 %, reflecting the compound probability of obtaining cloud-free satellite observations both immediately before and after grazing events. Excluded events exhibited similar duration characteristics (mean = 2.60 ± 2.60 days) compared to included events (mean = 2.86 ± 2.58 days), confirming that exclusions occurred due to temporal measurement constraints rather than systematic differences in event characteristics.

3.2.1. Pre- and post-grazing biomass agreement

Pre-grazing PB estimates exhibited near-perfect agreement between measurement systems ($R^2 = 0.966$, $p < 0.001$), with satellite estimates averaging 2934 ± 417 kg DM ha^{-1} compared to RPM measurements of 2913 ± 422 kg DM ha^{-1}. Bland-Altman analysis revealed minimal systematic bias (+20 kg DM ha^{-1}, 0.7 % of mean), though with moderate random variation (SD = 78 kg DM ha^{-1}, 95 % limits of agreement: -132 to +173 kg DM ha^{-1}). Both systems captured comparable PB ranges (RPM: 1632–3720 kg DM ha^{-1}; satellite: 1679–3720 kg DM ha^{-1}), spanning typical pre-grazing conditions from winter recovery to peak spring growth. Post-grazing PB demonstrated exceptional agreement ($R^2 = 0.998$, $p < 0.001$), with satellite estimates (2536 ± 459 kg DM ha^{-1}) nearly identical to RPM measurements (2539 ± 461 kg DM ha^{-1}). Systematic bias was negligible (-4 kg DM ha^{-1}, -0.1 % of mean) with minimal random variation (SD = 20 kg DM ha^{-1}, 95 % limits of agreement: -43 to +36 kg DM ha^{-1}). The substantially tighter agreement for post- versus pre-grazing measurements (SD reduction from 78 to 20 kg DM ha^{-1}) reflected reduced spatial heterogeneity following defoliation, where both measurement systems converged on more uniform residual sward heights.

3.2.2. Biomass removal and utilisation agreement

Pasture biomass removal estimates showed strong agreement between measurement systems ($R^2 = 0.922$, $p < 0.001$, RMSE = 83 kg DM ha^{-1}), with satellite-derived values (398 ± 286 kg DM ha^{-1}) marginally exceeding RPM measurements (374 ± 274 kg DM ha^{-1}). Bland-Altman analysis identified small but consistent positive bias (+24 kg DM ha^{-1}, 6.4 % of mean removal) with 95 % limits of agreement spanning -132 to +180 kg DM ha^{-1} (Fig. 6).

Farm-level correlations ranged from $R^2 = 0.659$ to $R^2 = 0.992$ (mean = 0.920), with RMSE varying from 14 to 230 kg DM ha^{-1} (mean = 65 kg DM ha^{-1}) across the 12 commercial operations (Fig. 7). Regional clustering within the farm-level scatter demonstrated consistent measurement agreement despite diverse management practices and environmental conditions, with all farms tracking close to the 1:1 agreement line. Agreement strength varied systematically with grazing intensity. Low-utilisation events (<10 % removal, n = 104) exhibited moderate correlation ($R^2 = 0.460$, RMSE = 77 kg DM ha^{-1}) with mean removal of 169 kg DM ha^{-1} (RPM) versus 193 kg DM ha^{-1} (satellite). Medium-utilisation events (10–20 %, n = 73) demonstrated stronger agreement ($R^2 = 0.764$, RMSE = 56 kg DM ha^{-1}) at 402 versus 423 kg DM ha^{-1}. High-utilisation events (>20 %, n = 41) maintained strong correlation ($R^2 = 0.719$) despite larger absolute errors (RMSE = 126 kg DM ha^{-1}) at mean removals of 844 versus 873 kg DM ha^{-1}. The positive bias (+24 to +29 kg DM ha^{-1}) remained consistent across utilisation categories, whilst correlation strengthened and proportional error declined as grazing intensity increased.

Pasture utilisation rate agreement paralleled removal patterns ($R^2 = 0.943$, $p < 0.001$, RMSE = 2.3 %), with satellite estimates (13.6 ± 9.3 %) closely matching RPM measurements (12.8 ± 9.1 %). Bland-Altman bias was minimal (+0.7 percentage points, 5.5 % of mean) with tight limits of agreement (-3.6 to +5.1 percentage points). Paddock-level pre- and post-grazing biomass measurements (Fig. 8) illustrated concordance between systems across representative farms from each region, where RPM and satellite estimates tracked closely despite varying seasonal conditions and grazing intensities (range: 1402–3720 kg DM ha^{-1}). The observed concordance reflects the training relationship between measurement systems, as satellite estimates were derived from a model trained on RPM ground truth data. Both measurement systems successfully captured the transition from pre-grazing peaks approaching 3000 kg DM ha^{-1} to post-grazing residuals near 1800 kg DM ha^{-1}, demonstrating consistent PB quantification across the full range of pasture utilisation scenarios.

4. Discussion

This study evaluated supervised (RF) and semi-supervised (OCSVM) ML approaches for automated GE detection and PU quantification from RPM measurements and interpolated Sentinel-2 satellite data across 12 commercial dairy farms in New South Wales. This work represents

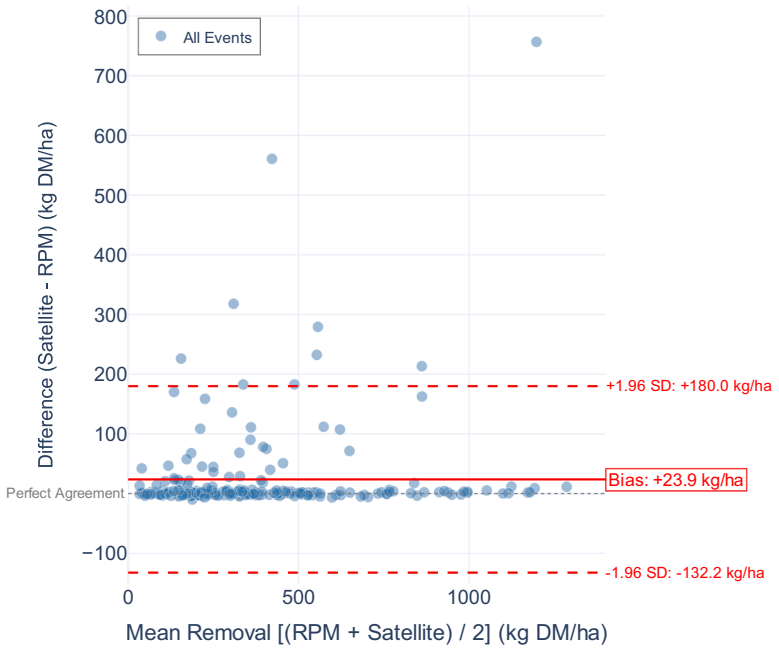


Fig. 6. Bland-Altman analysis of biomass removal agreement between measurement systems. Difference (Satellite - RPM) plotted against mean removal for 218 matched grazing events. Horizontal solid red line indicates bias (mean difference: +23.9 kg DM ha^{-1}); dashed red lines show 95 % limits of agreement (-132.2 to +180.0 kg DM ha^{-1}). Grey dotted line represents perfect agreement (difference = 0). Points colored by region: brown (Mid Coast), sea green (North Coast), steel blue (South Coast). Semi-transparent markers with slight jitter reveal density of overlapping observations. Clustering of points near zero difference indicates strong measurement agreement, with satellite estimates showing small positive bias across the removal range.

among the first systematic evaluations of automated GE detection specifically for P-BDS, addressing the need for scalable monitoring approaches in rotational grazing management. The findings reveal both the potential and limitations of these approaches for practical application across diverse P-BDS, particularly the challenges of temporal transferability when models are applied to years with environmental conditions differing from those in the training data.

4.1. *Grazing event detection performance and temporal transferability*

The within-year classification performance of RF for GE detection (Table 1) compares favourably to similar vegetation disturbance detection tasks in grassland systems. Studies on grassland mowing detection using comparable satellite data and ML approaches have reported F1-scores ranging from 0.61 to 0.94, depending on environmental conditions and temporal data availability [40,41]. The superior detection performance observed in this study likely reflects the distinct PB disappearance signatures associated with concentrated pasture intake in rotational grazing systems, where rapid complete utilisation of designated paddocks creates more pronounced temporal signals than the gradual PB removal in continuous or extensive grazing regimes. The consistency of RF performance across the three study regions (Fig. 7) suggests that the model successfully captured these spectral-temporal signatures of grazing disturbance despite regional variations in pasture composition, climate, and management practices.

Temporal transferability analyses revealed substantial performance degradation when models trained on Year 1 data were applied to Year 2 validation events (Fig. 3). This degradation aligns with findings from crop classification studies documenting reduced accuracy when models trained on single growing seasons encounter environmental variability in subsequent years [42]. Hoppe et al. [43] observed F1-score reductions from 82 % for single-day training to 61 % for full-season training, representing degradation of 10–21 % due to phenological variability between years. The 24.2 % average performance degradation observed for supervised models in this study exceeds these reported rates for crop classification tasks [44], highlighting the importance of model selection for temporal transferability in GE detection systems. While Random Forest achieved superior within-year detection performance, its reliance on year-specific training data limited cross-year generalisation. In contrast, the OCSVM semi-supervised framework demonstrated substantially better temporal robustness, outperforming supervised models by 7.6 % in Year 2 validation despite substantially lower absolute performance degradation. This pattern suggests that anomaly detection approaches are better suited for multi-year deployment without annual retraining. This heightened sensitivity to temporal shifts likely reflects the interaction between multiple dynamic factors including pasture species composition, seasonal growth patterns, grazing management timing, and year-to-year climatic variability. Filippelli et al. [45] further demonstrated that temporal cross-validation introduces systematic bias in model evaluation, as training and testing on the same year fails to account for interannual environmental variability. The primary driver of temporal performance degradation appears to be between-year environmental variability affecting pasture growth patterns and spectral responses. Phenological development of pasture species varies between years in response to temperature, rainfall, and soil moisture conditions. These changes alter the baseline spectral-temporal trajectories against which GE-induced PB reductions are detected, requiring model recalibration to maintain detection accuracy across multiple production years.

The observation that RPM-based models marginally outperformed MQ-based models in cross-year scenarios (average F1 difference of +0.044) provides insights into how temporal interpolation affects detection robustness when models are applied to subsequent years. Whilst MQ interpolation enables denser time series by filling temporal gaps between sparse RPM measurements, this process introduces smoothing inherent in this process that may reduce detection robustness under temporal environmental changes between years. Direct RPM measurements, despite lower temporal frequency, better preserve the abrupt biomass decline signals that occur during rotational grazing, particularly when measurements are collected shortly before and after grazing. This finding has implications for model training strategies, demonstrating that detection accuracy depends not solely on temporal

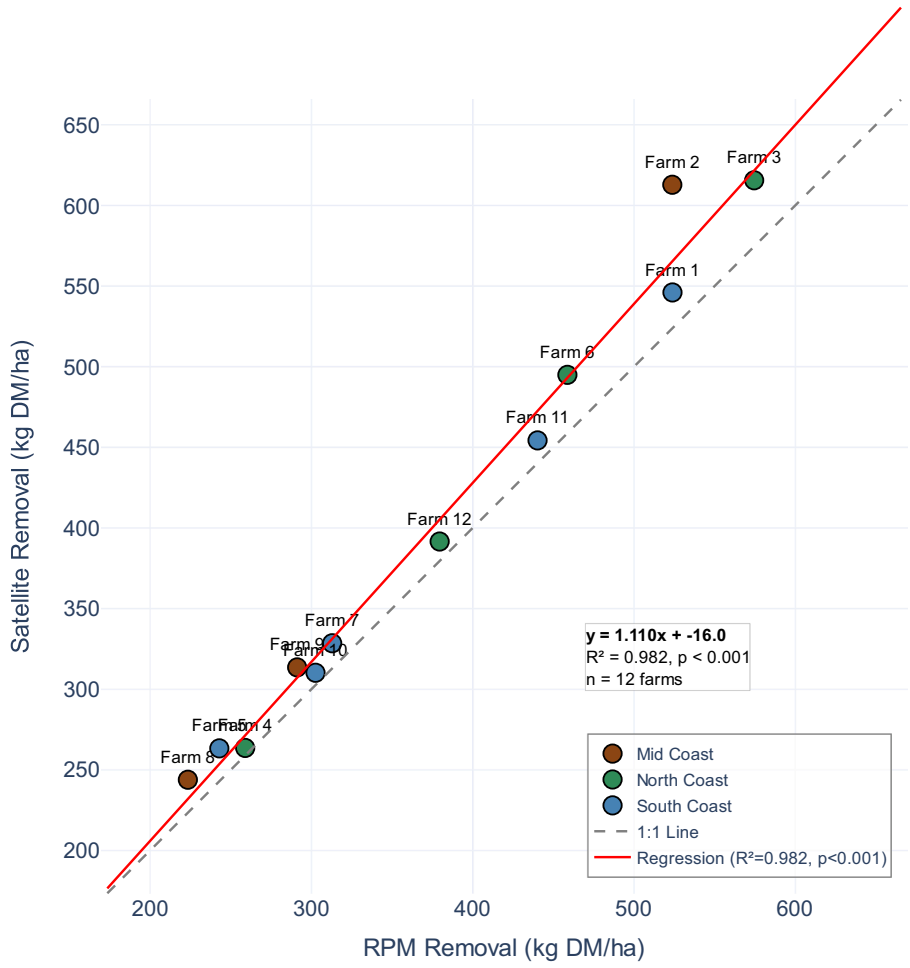


Fig. 7. Farm-level biomass removal agreement across measurement systems. Each point represents the mean removal estimate for one farm, calculated by averaging all GPS-confirmed grazing events within that farm (n = 12 farms, 218 total events). Points show RPM versus satellite measurements, colored by region: brown (Mid Coast, n = 3 farms), sea green (North Coast, n = 4 farms), steel blue (South Coast, n = 5 farms). Farm labels indicate individual operations (Farm 1–12). Grey dashed line represents perfect agreement (1:1); red solid line shows linear regression. Regional clustering demonstrates consistent measurement agreement across diverse management systems and environmental conditions, with satellite estimates tracking RPM measurements across the full range of grazing intensities.

resolution but also on how well the measurement approach captures the rapid PB disappearance events that occur during rotational grazing. However, the marginal nature of this advantage (+0.044 F1) suggests that interpolation remains valuable for improving within-year detection performance, with the trade-off between temporal density and signal preservation requiring careful consideration in practical applications.

The mechanism underlying this performance difference relates to how interpolation algorithms construct continuous time series from discrete observations. Multiquadric radial basis function interpolation generates smooth transitions between measured points by minimising curvature, which reduces abrupt variations in the resulting time series. In rotational grazing systems, genuine GE manifest as rapid biomass disappearance over 24 to 48 h, creating sharp temporal signals in direct measurements. The interpolation process necessarily dampens these sharp signals by distributing biomass changes across multiple interpolated points between actual measurements, making the characteristic rapid decline signature less pronounced. This smoothing effect reduces model sensitivity to abrupt grazing-induced changes whilst improving temporal coverage, explaining why RPM-based models maintained slightly higher cross-year detection performance despite lower temporal density.

The superior performance of RF compared to OCSVM for within-year GE detection (Table 1) reflects fundamental differences in how these approaches model the detection task. RF constructs ensemble decision boundaries from labelled training data, learning complex nonlinear relationships between spectral-temporal features and GE occurrence, enabling it explicitly to distinguish grazed from ungrazed observations based on supervised examples. This supervised learning paradigm excels when training and testing data share similar environmental conditions, as the model interpolates within the feature space defined by training samples. In contrast, OCSVM operates in a semi-supervised paradigm, learning a decision boundary around normal pasture growth patterns without explicit GE labels. This approach treats GE as anomalies rather than as a distinct class, making it more susceptible to extrapolation errors when environmental conditions in application years differ from the training year. The ensemble nature of RF provides additional robustness through averaging predictions across multiple decision trees; each trained on different bootstrap samples of the training data. This averaging reduces sensitivity to individual training samples whilst capturing diverse aspects of the GE detection signal. However, this same ensemble averaging can lead to overfitting when the training data inadequately represent the environmental variability encountered during operational deployment, explaining why RF performance degraded more substantially than OCSVM under cross-year validation. Meyer and Pebesma [46] characterised this challenge as the need to estimate the area of applicability for spatial prediction models, arguing that models should explicitly quantify when predictions extend beyond the environmental conditions represented in training data.

Despite its lower absolute performance, OCSVM nonetheless demonstrates the feasibility of semi-supervised approaches for GE detection

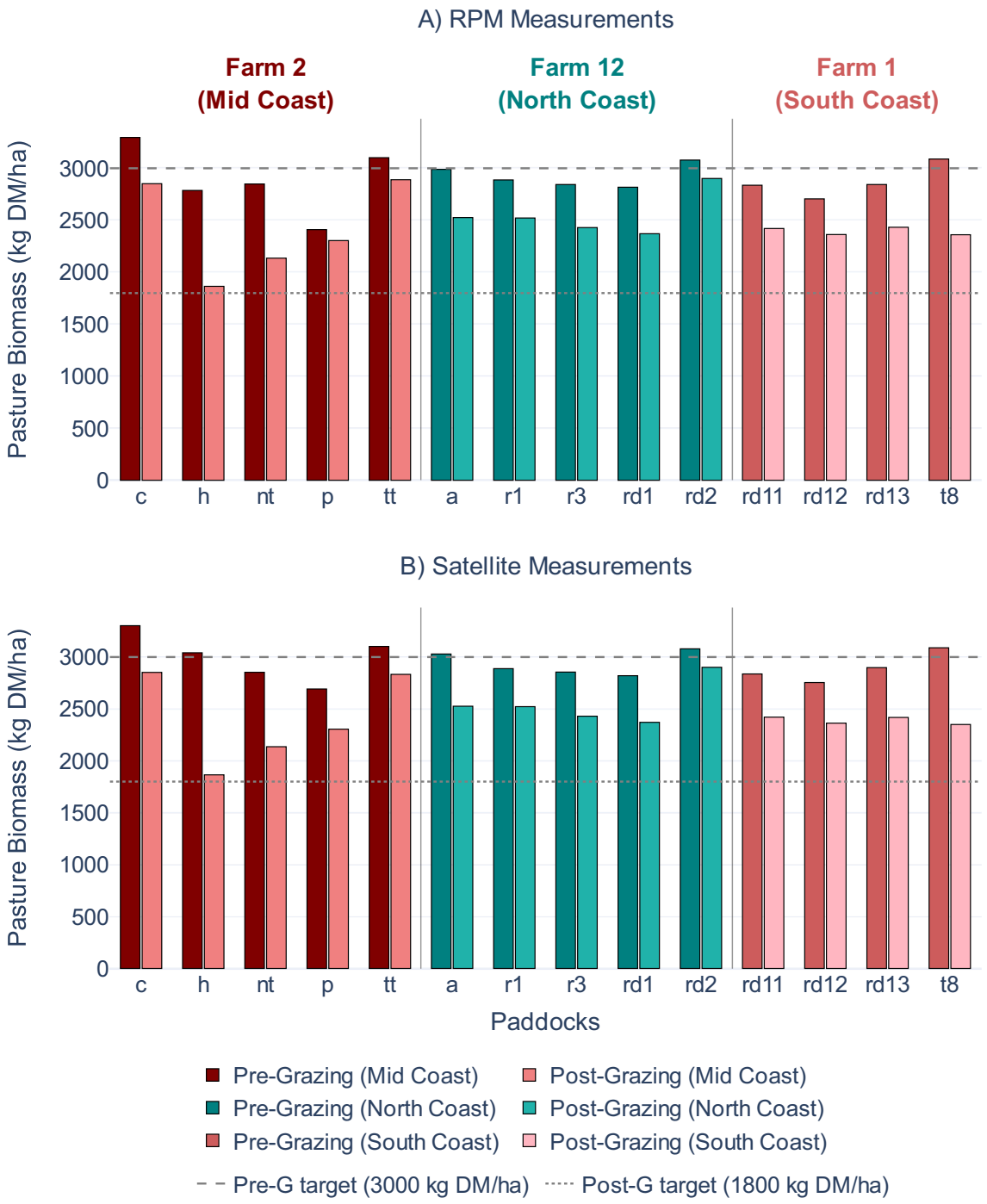


Fig. 8. Paddock-level pre-grazing (darker) and post-grazing (lighter) biomass from RPM (Panel A) and satellite (Panel B) measurements for representative farms from each region (Farm 2, Mid Coast; Farm 12, North Coast; Farm 1, South Coast). Horizontal dashed and dotted lines indicate pre- and post-grazing targets (3000 and 1800 kg DM ha^{-1}), respectively.

when labelled training data are limited or unavailable for specific years. The approach offers particular value in scenarios where obtaining contemporaneous RPM measurements is resource-intensive, as it requires training data from only a single reference year to establish baseline patterns. Recent advances in unsupervised adaptation techniques for remote sensing applications suggest potential pathways for improving cross-year transferability through transfer learning frameworks that explicitly model temporal shifts in feature distributions such as adversarial learning and self-training [47]. Future iterations of this approach could incorporate such techniques to enhance OCSVM performance, potentially combining limited labelled data from the target year with extensive historical data through transfer learning frameworks that explicitly model temporal environmental changes. To explore this direction, a preliminary hybrid scenario was evaluated wherein RF was retrained on Year 1 data augmented with increasing proportions of labelled Year 2 calibration data (10 %, 20 %, and 30 %). Results indicated that adding 10 % Year 2 calibration data (n = 87 observations) produced inferior performance relative to OCSVM (F1 = 0.676, Precision = 0.632, Recall = 0.726), whilst augmenting with 20 % (n = 174) yielded comparable performance (F1 = 0.693, Precision = 0.651, Recall = 0.742). The best hybrid scenario, incorporating 30 % Year 2 calibration data (n = 261), achieved marginal improvement over OCSVM (F1 =

0.708, Precision = 0.648, Recall = 0.781 vs OCSVM F1 = 0.692, Precision = 0.610, Recall = 0.799, +2.4 % relative improvement). These findings suggest that even modest Year 2 data augmentation can partially recover supervised model performance; however, the requirement for labelled Year 2 calibration data reduces the practical utility of RF approaches, as OCSVM achieves comparable temporal transferability without any target-year labels. This balance between ground-truth data requirements and detection accuracy represents an important consideration for the practical application of satellite-based grazing monitoring systems.

The regional consistency of RF performance (Fig. 7) suggests that within-year models can generalise effectively across spatially diverse farming environments, provided that training data adequately represent the range of pasture types, management intensities, and environmental conditions present in the target region. This finding supports the viability of developing regional or state-level GE detection systems trained on representative multi-farm datasets, rather than requiring farm-specific model training. However, the temporal transferability limitations of supervised models indicate that RF-based models would require annual retraining or updating with contemporary reference data to maintain classification accuracy across multiple years. In contrast, the superior cross-year performance of OCSVM demonstrates viability for multi-year deployment with reduced retraining requirements, though accepting lower absolute detection accuracy in exchange for temporal robustness.

4.2. Pasture utilisation agreement and measurement quality

Beyond GE detection, this study evaluated the agreement between satellite-derived PU estimates and independent RPM measurements across 218 GPS-confirmed GE. The observed agreement levels (Fig. 6) demonstrate that Sentinel-2 spectral indices can quantify PB removal with sufficient accuracy to support farm-level pasture management decisions, whilst revealing important measurement considerations affecting quantification reliability. The strong agreement between satellite-derived and RPM-based PU estimates for pre-grazing and post-grazing PB (Fig. 8) indicates that interpolated Sentinel-2 observations capture pasture states with high accuracy when temporal bracketing is adequate. The slightly lower agreement for PB removal calculations reflects the propagation of measurement uncertainties through the difference calculation, where small errors in both pre-grazing and post-grazing estimates compound. This pattern is consistent with error propagation principles and highlights a fundamental challenge in difference-based PB quantification approaches. Satellite-derived estimates must capture not just absolute PB levels but also the magnitude of rapid dairy cow pasture intake events, which introduces additional uncertainty compared to simple PB state estimation. Similar agreement patterns have been reported in other Sentinel-2 based pasture monitoring studies in Australian dairy systems, with reported R^2 values ranging from 0.60 to 0.94 depending on calibration approaches and environmental conditions [19].

The farm-level heterogeneity in agreement patterns (Fig. 8) provides insights into factors affecting PU quantification reliability. Farms achieving higher correlation values between satellite and RPM measurements typically exhibited more consistent grazing management practices, with regular measurement intervals and reliable temporal bracketing of GE. In contrast, farms with lower correlation values often experienced irregular measurement schedules or GE occurring outside the optimal temporal window for satellite observation. This variation shows that satellite-based PU quantification performs best within structured rotational grazing systems where management practices align with satellite revisit schedules, whilst becoming less reliable in more flexible grazing regimes. These findings align with observations from other satellite-based pasture validation studies, where measurement frequency and temporal alignment between ground and satellite observations strongly influence agreement levels [15,48].

Grazing intensity emerged as another factor influencing quantification accuracy, with higher-intensity GE (complete paddock utilisation) generally showing stronger agreement between measurement approaches than partial utilisation scenarios. Complete utilisation creates larger spectral changes more readily detected by Sentinel-2, whilst partial grazing generates subtler spectral signals that may approach the detection threshold of satellite instruments. This intensity-dependent accuracy has implications for system calibration, suggesting that satellite-based approaches may require adjustment factors accounting for expected grazing intensity levels within specific management conditions. The comparison between RPM measurements and satellite-derived PB estimates addresses a fundamental question about ground-truth data quality in remote sensing validation studies. Whilst RPM measurements are commonly treated as reference standards for pasture assessment, they introduce their own uncertainties through sampling variability, instrument calibration drift, operator technique, and conversion from compressed sward height to PB using site-specific calibration equations [49]. These measurement errors then propagate into both the training labels used for model development and the validation metrics used to assess satellite-derived estimates, creating potential circularity in accuracy assessment.

However, several factors support the validity of RPM-based validation in this scenario. First, RPM measurements were collected systematically across multiple transects per paddock reducing individual measurement variance through spatial averaging. Second, site-specific calibration relationships between RPM readings and PB were developed using destructive sampling techniques conducted across the full range of pasture conditions encountered in each region. Third, the consistency of agreement patterns across farms and regions (Fig. 8) suggests that RPM measurement errors are sufficiently controlled to reveal systematic differences in satellite-derived PU accuracy rather than being dominated by random RPM variance. This approach aligns with recommendations from recent pasture remote sensing validation studies emphasising the importance of systematic ground-truth collection protocols and site-specific calibrations [50,51].

The role of temporal interpolation in PU quantification presents a complex trade-off between measurement frequency and signal preservation. MQ interpolation enables PU estimation for GE occurring between sparse RPM measurement dates by constructing continuous PB trajectories from discrete observations. However, the interpolation process necessarily introduces smoothing that may obscure rapid PB changes characteristic of intensive grazing. The similar agreement levels observed for RPM-based and MQ-based PU estimates (Fig. 6) suggest that interpolation errors remain within acceptable bounds for farm applications, though the magnitude of these errors likely varies with measurement density, growth rate variability, and GE timing relative to adjacent measurements.

An important limitation in PU quantification accuracy stems from the assumption that all observed PB reduction represents dairy cow pasture intake, when in fact pasture trampling, senescence, and weather damage could also contribute to PB disappearance between measurements. RPM measurements integrate all sources of PB change, not solely intake, leading to potential overestimation of actual dairy cow pasture intake. Satellite-derived estimates inherit this limitation when validated against RPM measurements. Distinguishing pasture intake from other loss mechanisms would require additional data sources such as behaviour sensors or complementary remote sensing modalities that separate structural changes (trampling) from PB removal (intake).

Despite these measurement complexities, the demonstration that satellite-derived estimates achieve strong agreement with RPM measurements across diverse farm environments and grazing intensities provides confidence in the practical utility of remote sensing approaches for PU monitoring. The agreement levels observed in this study are consistent with R^2 values reported in other Sentinel-2 based pasture monitoring studies in Australian dairy systems [19,37].

4.3. Operational implications for pasture-based dairy systems

The demonstrated performance of automated GE detection and PU quantification presents practical opportunities for improving pasture management in commercial dairy operations. Current industry practice relies heavily on visual assessment or periodic RPM measurements, which are time-intensive and provide limited temporal coverage. The satellite-based approach demonstrated here enables continuous monitoring across entire farm areas without manual data collection, potentially reducing labour requirements whilst maintaining management decision accuracy. The choice between modelling approaches presents distinct implementation requirements. Farms adopting supervised models would need to maintain systematic RPM measurement programmes annually to generate training data for model recalibration, as the cross-year performance degradation suggests regular retraining may be necessary to maintain detection accuracy. Alternatively, OCSVM-based detection requires training data from only a single reference year to establish baseline pasture growth patterns, potentially enabling multi-year deployment without annual recalibration, though at the cost of reduced absolute detection accuracy (F1 = 0.692 vs. 0.878 for RF). The spatial coverage advantages of satellite monitoring are particularly relevant for larger dairy operations where manual measurement of all paddocks becomes impractical. The 10-metre resolution of Sentinel-2 imagery provides sufficient detail for paddock-scale management decisions in most commercial dairy systems. However, the temporal limitations imposed by cloud cover and the 5-day revisit frequency mean that satellite-based approaches would likely complement rather than completely replace ground-based measurements. Maintaining some level of RPM monitoring could enable validation of detection performance and capture of rapid biomass changes between satellite observations.

4.4. Limitations and future research directions

Four primary considerations emerge from this research that inform deployment strategies and guide future development priorities. First, temporal data alignment requirements, whilst validating the methodological approach on 218 events, limited validation comprehensiveness to 19.4 % of total GPS-confirmed GE. This limitation does not reflect methodological inadequacy but rather highlights the critical dependency of satellite-based PU monitoring on dense temporal observations capable of bracketing discrete GE. The 218 successfully validated events demonstrated strong agreement metrics (R^2 = 0.922–0.998), establishing proof-of-concept that the integrated detection-quantification framework functions effectively when adequate temporal alignment exists. However, applications at commercial scale requires substantially improved temporal coverage to capture the full spectrum of management scenarios encountered across diverse farming operations.

The fundamental challenge lies in capturing rapid PB changes characteristic of rotational grazing, where paddock-level PB can decline by 500–1500 kg DM ha^{-1} within 24–48 h during active grazing periods. Current satellite revisit frequencies (Sentinel-2: 5 days nominal) combined with cloud contamination constraints create temporal gaps that miss these critical PB disappearance signals, preventing accurate pre-grazing peak identification and post-grazing residual quantification for the majority of GE. The compound probability of obtaining cloud-free observations both immediately before and shortly after grazing explains why 80.6 % of events lacked complete temporal bracketing, with missing observations exhibiting spatial clustering in coastal high-rainfall regions and temporal clustering during winter cloud-prone periods (inclusion rate: 9.7 % vs 24.1 % in autumn). However, even optimal seasonal conditions (autumn) demonstrated 75.9 % exclusion rates, indicating that the limitation reflects fundamental satellite revisit frequency constraints compounded by cloud interference rather than seasonal weather patterns alone. Despite these substantial temporal limitations, satellite-based GE detection remains appropriate for specific applications. The 218 successfully validated events demonstrate that when adequate temporal alignment exists, the approach achieves reliable detection accuracy and PU quantification. Current satellite systems are most appropriate for: (1) retrospective analysis of grazing patterns where timing precision is less critical, (2) farms with structured rotational systems aligned with satellite revisit schedules, and (3) supplementary validation of ground-based measurements rather than complete replacement. However, use as a primary real-time monitoring tool requires the temporal density improvements discussed below to capture the majority of GE across diverse management scenarios.

Future research must prioritise achieving daily or near-daily biomass observation density to overcome these constraints. Multi-sensor fusion approaches integrating Sentinel-1 C-band SAR (6-day revisit, all-weather capability) with Sentinel-2 optical imagery represent the most promising pathway for achieving this temporal density. SAR backscatter responds to vegetation structure and moisture content, providing PB-sensitive measurements during cloudy conditions when optical sensors fail [52]. Successful fusion of optical and SAR time series for grassland mowing detection has demonstrated F1 scores of 89 % [40], suggesting that similar integration could substantially improve GE detection reliability whilst enabling continuous PU monitoring regardless of atmospheric conditions. However, SAR-optical fusion for rotational grazing in P-BDS requires addressing specific challenges including cross-sensor calibration for PB estimation, temporal interpolation algorithms that preserve abrupt intake signals and validation frameworks accounting for different measurement principles.

Dense temporal observations offer multiple advantages beyond simply increasing validation sample size. First, daily PB estimates have potential to enable precise identification of pre-grazing peaks and post-grazing minima at paddock scale, reducing temporal alignment uncertainty that currently constrains removal accuracy. Second, continuous time series will facilitate detection of gradual versus rapid PB changes, improving discrimination between grazing-induced intake and senescence-driven declines that currently confound detection during extended intervals between measurements. Third, higher temporal density improves interpolation quality by shortening the intervals between actual measurements, enabling interpolation across 1–2 day gaps rather than 7–14 day periods typical of weekly to fortnightly RPM collection schedules that was used in this present study, thereby minimising error accumulation and reducing propagation of systematic measurement bias from RPM calibration equations through MQ-derived estimates. Fourth, paddock-level PB trajectories spanning complete grazing cycles enable validation of growth rate predictions and residual target compliance monitoring that single-timepoint measurements cannot support. These advantages collectively transform satellite monitoring from retrospective validation tools to prospective decision support systems capable of informing real-time grazing management.

Additionally, the semi-supervised OCSVM framework, whilst demonstrating superior temporal transferability, achieved moderate absolute performance (F1 = 0.692) that may benefit from refinement for autonomous use. Hybrid approaches combining OCSVM anomaly detection for initial event flagging with supervised verification models could potentially achieve both high transferability and improved precision, reducing false positive rates whilst maintaining robust recall across diverse temporal conditions. Development of confidence scoring mechanisms that quantify detection uncertainty would enable operators to prioritise verification efforts on ambiguous classifications whilst automating high-confidence detections. Furthermore, the study focused on coastal regions in New South Wales, requiring validation across diverse pasture species, climatic zones and management intensities to establish broader applicability. The observed farm-level performance heterogeneity (F1 range: 0.500–0.815) and its moderate inverse relationship with stocking rate (r = -0.504) suggest detection accuracy depends on complex interactions between grazing management intensity, pasture growth dynamics, and satellite observation frequency. Higher

stocking rates may reduce detection performance by accelerating biomass depletion, thereby narrowing the temporal window between pre-grazing and post-grazing satellite observations. This finding has implications: satellite-based grazing detection systems may require farm-specific calibration or adaptive detection thresholds to account for varying management intensities. Controlled experiments manipulating stocking rate, rotation frequency and spelling duration whilst maintaining consistent monitoring protocols would isolate how management decisions influence detection reliability, informing best-practice guidelines for satellite-based grazing monitoring.

Finally, validation relied on GPS ground truth from 3 collared cows per farm, which provided operational feasibility whilst introducing potential limitations in capturing complete herd behaviour and sub-paddock utilisation heterogeneity. Integration with complementary ground truth sources including automated gate monitoring systems, virtual fencing platforms and farm management software could strengthen validation robustness whilst enabling detection of partial grazing scenarios where livestock access only portions of designated paddocks. Future research should explore optimal automated ground truth collection strategies balancing data quality, operational costs and commercial farming constraints.

5. Conclusion

This research established that machine learning based frameworks can successfully automate grazing event detection in pasture-based dairy systems, with Random Forest achieving within-year F1 scores of 0.878 and One-Class Support Vector Machine (OCSVM) demonstrating superior temporal transferability with cross-year F1 scores of 0.692. The contrasting performance patterns between supervised and anomaly detection approaches reveal fundamental trade-offs between absolute accuracy and temporal generalisation, with OCSVM semi-supervised learning paradigm proving more robust to interannual variability in pasture growth dynamics and management practices. These findings establish that satellite-based grazing event detection is operationally viable for commercial deployment when appropriate modelling strategies are employed. Strong pasture utilisation agreement metrics (R^2 = 0.922–0.998) validated that satellite systems can accurately quantify pre-grazing pasture biomass, post-grazing residuals and removal amounts when adequate temporal alignment exists, supporting their use as reliable data sources for automated monitoring frameworks. The integration of automated detection with independent pasture biomass quantification addresses critical gaps in current pasture monitoring systems, with potential to enable data-driven stocking rate adjustments and residual target compliance assessment without manual intervention. However, model deployment requires substantially improved temporal observation density to overcome the 80.6 % event exclusion rate resulting from cloud contamination and satellite revisit constraints. Multi-sensor fusion integrating Sentinel-1 synthetic aperture radar (SAR) with Sentinel-2 optical imagery represents the most viable pathway for achieving daily pasture biomass observations, enabling interpolation across 1–2 day intervals rather than 7–14 day gaps typical of current rising plate meter collection schedules. Dense time series would transform satellite monitoring from retrospective validation tools to prospective decision support systems capable of informing real-time grazing management in commercial dairy operations.

Funding statement

This research was funded by the Dairy UP program, a collaborative RD&E program for New South Wales, Australia (www.dairyup.com.au) through the academic scholarship awarded to Blessing N. Azubuike.

Ethics statement

If Yes, please provide your text here:

This study was conducted in accordance with the ethical standards outlined by the relevant institutional and national guidelines. The research involving animals was reviewed and approved by The University of Sydney's Animal Ethics Committee (Protocol Number: 2019/1608). All procedures performed in this study adhered to ethical principles, ensuring the welfare of the animals involved. The study complied with the Australian Code for the Care and Use of Animals for Scientific Purposes, 8th Edition (2013).

CRediT authorship contribution statement

Blessing Nnenna Azubuike: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Chlingaryan:** Writing – review & editing, Supervision, Conceptualization. **Martin Correa-Luna:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Cameron E.F. Clark:** Writing – review & editing, Supervision, Conceptualization. **Sergio C. Garcia:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the research scholarship granted by the DairyUP project at the University of Sydney.

Data availability

Data will be made available on request.

References

[1] P. Beukes, S. McCarthy, C. Wims, P. Gregorini, A. Romera, Regular estimates of herbage mass can improve profitability of pasture-based dairy systems, Anim Prod Sci 59 (2018) 359–367, https://doi.org/10.1071/AN17166.
[2] J.R. Roche, D.P. Berry, A.M. Bryant, C.R. Burke, S.T. Butler, P.G. Dillon, D. J. Donaghy, B. Horan, K.A. Macdonald, K.L. Macmillan, A 100-year review: A century of change in temperate grazing dairy systems, J. Dairy. Sci. 100 (12) (2017) 10189–10233, https://doi.org/10.3168/jds.2017-13182.
[3] D. Chapman, J. Bryant, M. Olayemi, G. Edwards, B. Thorrold, W. McMillan, G. Kerr, G. Judson, T. Cookson, A. Moorhead, M. Norriss, An economically based evaluation index for perennial and short-term ryegrasses in New Zealand dairy farm systems, Grass. Forage Sci. 72 (2016), https://doi.org/10.1111/gfs.12213 n/a-n/a.
[4] D.W. Bailey, M.G. Trotter, C.W. Knight, M.G. Thomas, Use of GPS tracking collars and accelerometers for rangeland livestock production research, Transl. Anim. Sci. 2 (1) (2018) 81–88, https://doi.org/10.1093/tas/txx006.
[5] A. Herlin, E. Brunberg, J. Hultgren, N. Högberg, A. Rydberg, A. Skarin, Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture, Animals 11 (3) (2021) 829. https://www.mdpi.com/2076-2615/11/3/829.
[6] L. Turner, C. Udal, B. Larson, S. Shearer, Monitoring cattle behavior and pasture use with GPS and GIS, Can. J. Anim. Sci. Can. J. Anim. Sci. 80 (2000) 405–413, https://doi.org/10.4141/A99-093.
[7] J. Barwick, D.W. Lamb, R. Dobos, M. Welch, M. Trotter, Categorising sheep activity using a tri-axial accelerometer, Comput. Electron. Agric. 145 (2018) 289–297, https://doi.org/10.1016/j.compag.2018.01.007.
[8] J. Cabezas, R. Yubero, B. Visitación, J. Navarro-García, M.J. Algar, E.L. Cano, F. Ortega, Analysis of accelerometer and GPS data for cattle behaviour identification and anomalous events detection, Entropy 24 (3) (2022) 336, https://doi.org/10.3390/e24030336.
[9] L. Riaboff, S. Poggi, A. Madouasse, S. Couvreur, S. Aubin, N. Bédère, E. Goumand, A. Chauvin, G. Plantier, Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data, Comput. Electron. Agric. 169 (2020) 105179, https://doi.org/10.1016/j.compag.2019.105179.
[10] J.A. Vázquez Diosdado, Z.E. Barker, H.R. Hodges, J.R. Amory, D.P. Croft, N.J. Bell, E.A. Codling, Classification of behaviour in housed dairy cows using an

accelerometer-based activity monitoring system, Anim. Biotelemetry. 3 (1) (2015) 15, https://doi.org/10.1186/s40317-015-0045-8.
[11] D.A. McGranahan, B. Geaumont, J.W. Spiess, Assessment of a livestock GPS collar based on an open-source datalogger informs best practices for logging intensity, Ecol. Evol. 8 (11) (2018) 5649–5660, https://doi.org/10.1002/ece3.4094.
[12] M. Parlato, F. Valenti, S. Porto, GIS-based methodology for tracking the grazing cattle site use, Heliyon 10 (2024) e33166, https://doi.org/10.1016/j.heliyon.2024.e33166.
[13] J.M. Chapa, B. Pichlbauer, M. Bobal, C. Guse, M. Drillich, M. Iwersen, Field evaluation of a rising plate meter to estimate herbage mass in Austrian pastures, Sensors 23 (17) (2023) 7477, https://doi.org/10.3390/s23177477.
[14] M. Sanderson, C.A. Rotz, S. Fultz, E. Rayburn, Estimating forage mass with a commercial capacitance meter, rising plate meter, and pasture ruler, Agron. J. 93 (2001), https://doi.org/10.2134/agronj2001.1281.
[15] D. De Rosa, B. Basso, M. Fasiolo, J. Friedl, B. Fulkerson, P. Grace, D. Rowlings, Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery, Comput. Electron. Agric. 180 (2021), https://doi.org/10.1016/j.compag.2020.105880.
[16] J. Gargiulo, C. Clark, N. Lyons, G. de Veyrac, P. Beale, S. Garcia, Spatial and temporal pasture biomass estimation integrating electronic plate meter, planet CubeSats and Sentinel-2 satellite data, Remote Sens. 12 (19) (2020) 3222, https://doi.org/10.3390/rs12193222.
[17] B.J. Heins, G.M. Pereira, K.T. Sharpe, Precision technologies to improve dairy grazing systems* *presented as part of the Joint ADSA midwest branch/forages and pastures symposium: grazing to improve profitability of midwest dairy farms held at the ADSA annual meeting, June 2022, JDS. Commun. 4 (4) (2023) 318–323, https://doi.org/10.3168/jdsc.2022-0308.
[18] A. Michez, P. Lejeune, S. Bauwens, A.A.L. Herinaina, Y. Blaise, E. Castro Muñoz, F. Lebeau, J. Bindelle, Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system, Remote Sens. 11 (5) (2019) 473. https://www.mdpi.com/2072-4292/11/5/473.
[19] Y. Chen, J. Guerschman, Y. Shendryk, D. Henry, M.T. Harrison, Estimating pasture biomass using sentinel-2 imagery and machine learni ng, Remote Sens. 13 (4) (2021) 603, https://doi.org/10.3390/rs13040603.
[20] M.G. Ogungbuyi, J. Guerschman, A.M. Fischer, R.A. Crabbe, I. Ara, C. Mohammed, P. Scarth, P. Tickle, J. Whitehead, M.T. Harrison, Improvement of pasture biomass modelling using high-resolution satelli te imagery and machine learning, J. Environ. Manag. 356 (2024) 120564, https://doi.org/10.1016/j.jenvman.2024.120564.
[21] M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: a meta-review, Remote Sens. Environ. 236 (2020) 111402, https://doi.org/10.1016/j.rse.2019.111402.
[22] B.N. Azubuike, A. Chlingaryan, M. Correa-Luna, C.E.F. Clark, S.C. Garcia, Data augmentation and interpolation improves machine learning-based pasture biomass estimation from sentinel-2 imagery, Remote Sens. 17 (23) (2025) 3787, https://doi.org/10.3390/rs17233787.
[23] I. Ali, F. Cawkwell, E. Dwyer, S. Green, Modeling managed grassland biomass estimation by using multitemporal R emote sensing data—a machine learning approach, IEEE J. Sel. Top. Appl. Earth Obs. Remo Te Sens. 10 (7) (2017) 3254–3264, https://doi.org/10.1109/jstars.2016.2561618.
[24] A. Edirisinghe, D. Clark, D. Waugh, Spatio-temporal modelling of biomass of intensively grazed perennial d airy pastures using multispectral remote sensing, Int. J. Appl. Earth. Obs. Geoinf. 16 (2012) 5–16, https://doi.org/10.1016/j.jag.2011.11.006.
[25] C. Stumpe, J. Leukel, T. Zimpel, Prediction of pasture yield using machine learning-based optical sensing: a systematic review, Precis. Agric. 25 (2023) 1–30, https://doi.org/10.1007/s11119-023-10079-9.
[26] O. Dubovik, G.L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, Z. Li, Grand challenges in satellite remote sensing [specialty grand challenge], Front. Remote Sens. (2021), https://doi.org/10.3389/frsen.2021.619818. *Volume 2 - 2021*.
[27] M.C. Hott, R.G. Andrade, P.S. D'Oliveira, M.B. Motta, W.S.D.D. Rocha, W.C.P. D. Magalhães Junior, Remote sensing applied to pasture monitoring: A review, Revista de Gestão RGSA 18 (7) (2024) e08355, https://doi.org/10.24857/rgsa.v18n7-161.
[28] H. Liu, P. Guo, J. Liu, R. Liu, T. Tong, An extension of multiquadric method based on trend analysis for surface construction, IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens. 16 (2023) 3435–3441, https://doi.org/10.1109/JSTARS.2023.3247598.
[29] Fabre Ferber, F., Gay, D., Soulié, J.C., Diatta, J., & Maillard, O.-A. (2025). *Kriging and Gaussian process interpolation for georeferenced data augmentation*. https://doi.org/10.48550/arXiv.2501.07183.
[30] W.A. Nuss, D.W. Titley, Use of multiquadric interpolation for meteorological objective analysis, Mon. Weather. Rev. 122 (1994) 1611–1631, https://doi.org/10.1175/1520-0493(1994)122<1611:UOMIFM>2.0.CO;2.
[31] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[32] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (2016) 785–794, https://doi.org/10.1145/2939672.2939785.
[33] R. García, J. Aguilar, A. Pinto, M. Toro, A systematic literature review on the use of machine learning in precision livestock farming, Comput. Electron. Agric. 179 (2020), https://doi.org/10.1016/j.compag.2020.105826.
[34] P. Shine, M.D. Murphy, Over 20 years of machine learning applications on dairy farms: a comprehensive mapping study, Sensors 22 (1) (2022) 52. https://www.mdpi.com/1424-8220/22/1/52.
[35] E. Fogarty, D. Swain, G. Cronin, L. Moraes, D. Bailey, M. Trotter, Potential for autonomous detection of lambing using global navigation satellite system technology, Anim. Prod. Sci. 60 (2020) 1217–1226, https://doi.org/10.1071/AN18654.
[36] Digital Matter. (2025). Yabby edge cellular GPS tracker. Retrieved 18 December 2025 from https://www.digitalmatter.com/devices/yabby-edge-cellular/.
[37] M. Correa-Luna, J. Gargiulo, P. Beale, D. Deane, J. Leonard, J. Hack, Z. Geldof, C. Wilson, S. Garcia, Accounting for minimum data required to train a machine learning model to accurately monitor Australian dairy pastures using remote sensing, Sci. Rep. 14 (1) (2024), https://doi.org/10.1038/s41598-024-68094-3.
[38] Z. Zhu, C.E. Woodcock, Object-based cloud and cloud shadow detection in landsat imagery, Remote Sens. Environ. 118 (2012) 83–94, https://doi.org/10.1016/j.rse.2011.10.028.
[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, G. Louppe, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2012), https://doi.org/10.48550/arXiv.1201.0490.
[40] A.-K. Holtgrave, F. Lobert, S. Erasmi, N. Röder, B. Kleinschmit, Grassland mowing event detection using combined optical, SAR, and weather time series, Remote Sens. Environ. 295 (2023) 113680, https://doi.org/10.1016/j.rse.2023.113680.
[41] M. Schwieder, M. Wesemeyer, D. Frantz, K. Pfoch, S. Erasmi, J. Pickert, C. Nendel, P. Hostert, Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series, Remote Sens. Environ. 269 (2022) 112795, https://doi.org/10.1016/j.rse.2021.112795.
[42] T. Rusňák, T. Kasanický, P. Malík, J. Mojžiš, J. Zelenka, M. Sviček, D. Abrahám, A. Halabuk, Crop mapping without labels: investigating temporal and spatial transferability of Crop classification models using a 5-year Sentinel-2 series and machine learning, Remote Sens. 15 (13) (2023) 3414.
[43] H. Hoppe, P. Dietrich, P. Marzahn, T. Weiß, C. Nitzsche, U. Freiherr von Lukas, T. Wengerek, E. Borg, Transferability of machine learning models for crop classification in remote sensing imagery using a new test methodology: A study on phenological, temporal, and spatial influences, Remote Sens. 16 (9) (2024) 1493.
[44] C. Pelletier, S. Valero, J. Inglada, N. Champion, G. Dedieu, Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas, Remote Sens. Environ. 187 (2016) 156–168, https://doi.org/10.1016/j.rse.2016.10.010.
[45] S. Filippelli, K. Schleeweis, M. Nelson, P. Fekety, J. Vogeler, Testing temporal transferability of remote sensing models for large area monitoring, Sci. Remote Sens. 9 (2024) 100119, https://doi.org/10.1016/j.srs.2024.100119.
[46] H. Meyer, E. Pebesma, Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods Ecol. Evol. 12 (9) (2021) 1620–1633, https://doi.org/10.1111/2041-210X.13650.
[47] S. Luo, L. Ma, X. Yang, D. Luo, Q. Du, Self-Training-Based Unsupervised Domain Adaptation for Object Detection in Remote Sensing Imagery, IEEE Transactions on Geoscience and Remote Sensing, PP (2024) 1, https://doi.org/10.1109/TGRS.2024.3457789, 1.
[48] J.R. Insua, S.A. Utsumi, B. Basso, Estimation of spatial and temporal variability of pasture growth and d igestibility in grazing rotations coupling unmanned aerial vehicle (UA V) with crop simulation models, PLoS. One 14 (3) (2019) e0212773, https://doi.org/10.1371/journal.pone.0212773.
[49] D.J. Murphy, B. O' Brien, D. Hennessy, M. Hurley, M.D. Murphy, Evaluation of the precision of the rising plate meter for measuring compressed sward height on heterogeneous grassland swards, Precis. Agric. 22 (3) (2021) 922–946, https://doi.org/10.1007/s11119-020-09765-9.
[50] I. Ali, F. Cawkwell, N. Dwyer, B. Barrett, S. Green, Satellite remote sensing of grasslands: from observation to management–a review, J. Plant Ecol. 9 (2016), https://doi.org/10.1093/jpe/rtw005.
[51] J.I. Gargiulo, N.A. Lyons, F. Masia, P. Beale, J.R. Insua, M. Correa-Luna, S. C. Garcia, Comparison of ground-based, unmanned aerial vehicles and satellite rem ote sensing technologies for monitoring pasture biomass on dairy farms, Remote Sens. 15 (11) (2023) 2752, https://doi.org/10.3390/rs15112752.
[52] A. Mercier, J. Betbeder, F. Rumiano, J. Baudry, V. Gond, L. Blanc, C. Bourgoin, G. Cornu, C. Ciudad, M. Marchamalo, R. Poccard-Chapuis, L. Hubert-Moy, Evaluation of sentinel-1 and 2 time series for land cover classification of forest–Agriculture mosaics in temperate and tropical landscapes, Remote Sens. 11 (8) (2019) 979.

CHAPTER 7

Transfer learning and stacking ensembles for biomass estimation from smartphone imagery in pasture-based dairy systems

What if a simple smartphone photo could reveal pasture biomass instantly? Satellite imagery suffers from cloud cover, expensive sensors remain impractical, and manual rising plate meters demand time intensive labour. This chapter explores a practical ground-based alternative using everyday smartphone imagery. We evaluate deep learning methodologies including transfer learning and stacking ensembles combined with weather data to achieve optimal predictive accuracy for pasture biomass estimation across diverse commercial dairy farm environments. The complete biomass prediction implementation, including all experiments is available in the repository (Chapter_7_Biomass_Prediction_from_images_notebook.ipynb). Key code excerpts are presented in Appendix E.

Journal of Agriculture and Food Research (2026) 28, 102923

PUBLISHED MANUSCRIPT

The published version of this manuscript is included on the following page.

Journal of Agriculture and Food Research 28 (2026) 102923

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Transfer learning and stacking ensembles for biomass estimation from smartphone imagery in pasture-based dairy systems

Blessing Nnenna Azubuike [a,d,*], Anna Chlingaryan [b,d], Martin Correa-Luna [d], Cameron E.F. Clark [c,d], Sergio C. Garcia [a,d]

[a] Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW, 2570, Australia
[b] Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Camden, NSW, 2570, Australia
[c] Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW, 2650, Australia
[d] Dairy UP Program, Camden, NSW, 2570, Australia

ARTICLE INFO

Keywords:
Deep learning
Ensemble learning
Pasture biomass
Spatial generalisation
Transfer learning

ABSTRACT

Accurate estimation of pasture biomass is essential for optimising feed management in pasture-based dairy systems. Traditional methods remain labour-intensive and impractical for frequent monitoring. Remote sensing via satellites or unmanned aerial vehicles offers alternatives but is constrained by cloud interference and infrastructure costs. This study investigates smartphone imagery as a scalable solution, evaluating 21 machine learning configurations for predicting pasture biomass across 15 commercial dairy farms in New South Wales, Australia. Ground-level red-green-blue images (n = 3291) were paired with rising plate meter measurements and weather data to develop estimation models using transfer learning, ensemble methods, and multi-architecture feature integration. Under random within-dataset partitioning, stacking ensemble combining pre-trained convolutional neural network features with hand-crafted vegetation indices and weather variables achieved R^2 = 0.561, MAE = 351 kg DM ha^{-1}, and RMSE = 437 kg DM ha^{-1}, representing a 48% improvement over baseline hand-crafted features. Transfer learning using frozen ImageNet features outperformed deep learning trained from scratch. Multi-architecture integration captured complementary information, with DenseNet121 features dominating importance rankings. Weather data integration provided 5.9% performance improvement. Seasonal performance varied substantially, with spring achieving highest accuracy (R^2 = 0.701, MAE = 308 kg DM ha^{-1}) during peak ryegrass growth and winter showing reduced performance (R^2 = 0.456, MAE = 375 kg DM ha^{-1}). However, leave-one-farm-out cross-validation revealed severe spatial generalisation constraints (mean R^2 = −0.009, MAE = 399 kg DM ha^{-1}, RMSE = 488 kg DM ha^{-1}), exposing strong farm-specific dependencies that limit predictive accuracy for unseen farm environments. These findings indicate that the proposed approach performs best when target farms contribute to the training dataset; generalisation to completely new farm environments will require either substantially expanded training datasets or farm-specific adaptation strategies.

1. Introduction

Accurate quantification of pasture biomass (PB) represents a fundamental requirement for efficient grazing management in pasture-based dairy systems (P-BDS). Optimal pasture allocation decisions depend critically on precise estimates of available forage, yet traditional measurement methods like the rising plate meter (RPM) remain labour-intensive, time-consuming, and subject to considerable operator variability [1–4]. The RPM has emerged as the industry standard for ground-based PB assessment, providing reasonable accuracy through mechanical measurement of pasture height and density. However, this approach requires extensive manual traversing of paddocks to obtain representative samples, limiting its practical application for frequent monitoring across large farm areas [1,5]. The imperative for more efficient, scalable, and automated PB estimation methods has driven growing interest in image-based monitoring approaches that leverage advances in computer vision and machine learning (ML).

Recent progress in deep learning (DL) and computer vision has

* Corresponding author. Dairy Science Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Camden, NSW, 2570, Australia.
E-mail address: blessing.azubuike@sydney.edu.au (B.N. Azubuike).

https://doi.org/10.1016/j.jafr.2026.102923
Received 2 January 2026; Received in revised form 22 February 2026; Accepted 10 April 2026
Available online 11 April 2026

created unprecedented opportunities for automated agricultural monitoring through image-based analysis [6–8]. Convolutional neural networks (CNNs) have demonstrated remarkable capabilities in extracting hierarchical visual features from imagery, enabling accurate prediction of crop characteristics including biomass, yield, and quality parameters [9–11]. Transfer learning (TL) approaches, which leverage knowledge acquired from large-scale image datasets to initialise network weights, have proven particularly effective for agricultural applications where training data may be limited [12–14]. Pre-trained architectures such as Residual Network 50 (ResNet50) and Visual Geometry Group 16 (VGG16) have become foundational tools for agricultural computer vision applications, demonstrating effectiveness across diverse tasks including plant disease detection, crop classification, and biomass estimation [15–21].

Despite these technological advances, PB estimation from ground-level imagery presents distinct challenges. Pasture canopies exhibit substantial spatial heterogeneity at multiple scales, with complex three-dimensional structure arising from varied plant architectures, growth stages, and species composition (Harmoney et al., 1997). Temporal dynamics introduce additional complexity, as pasture appearance varies markedly across seasons and in response to environmental conditions including temperature, rainfall, and solar radiation [22,23]. Furthermore, different grass species demonstrate divergent growth patterns and visual characteristics, with cool-season ryegrass and warm-season kikuyu exhibiting fundamentally different responses to environmental drivers [24,25].

Existing ground-based biomass estimation studies have explored diverse approaches ranging from hand-crafted feature engineering to DL architectures, yet systematic comparisons of their relative merits for pasture applications remain limited [26–32]. Hand-crafted features based on colour indices, texture measures, and geometric properties have demonstrated moderate success in grassland biomass prediction but require substantial domain expertise for feature design and may not adequately capture complex visual patterns [33–35]. Deep learning methods can automatically learn hierarchical representations from raw imagery but typically require extensive training data. [8,36,37]. The potential benefits of combining these complementary approaches through ensemble methods remain largely unexplored in ground-level pasture monitoring contexts. Recent studies have demonstrated that ensemble ML methods including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Gradient Boosting Machine (GBM) can achieve superior performance compared to individual models [38–40], whilst stacking approaches with meta-learning have shown promise for integrating heterogeneous base learners [41]. However, these ensemble strategies have not been systematically evaluated for ground-level pasture imagery. A further critical and underexplored challenge is spatial generalisation: whether models trained on data from known farms can reliably predict biomass at new farm environments not represented during training, a prerequisite for practical adoption at scale.

Ground-level RGB imagery encodes PB information through several visual signals. Canopy greenness, as captured by visible-band vegetation indices such as the Excess Green Index and the Normalised Green-Red Difference Index, correlates with chlorophyll content and leaf area, both of which scale with above-ground dry matter accumulation in temperate grasses [31]. Canopy texture, derived from spatial patterns in pixel intensity, reflects sward density and leaf arrangement, providing complementary structural information to colour-based indices [30]. However, RGB imagery has recognised limitations for PB estimation. Prostrate species such as kikuyu present low canopy profiles that are poorly differentiated by colour alone, and RGB sensors cannot directly capture canopy height, which is a strong driver of dry matter content in erect species like ryegrass [42]. Spectral saturation at high PB levels and variable illumination conditions further constrain the sensitivity of visible-band indices, establishing a practical ceiling on prediction accuracy from RGB imagery alone regardless of model complexity. Additionally, the contribution of environmental variables such as weather data to image-based PB prediction has not been rigorously quantified, despite evidence from satellite-based studies that they significantly influence prediction accuracy [43,44].

Critical questions remain unanswered regarding optimal model architecture for operational PB estimation. It remains unclear whether TL with pre-trained CNNs can outperform DL trained from scratch for agricultural imagery with moderate sample sizes. The complementarity of features from multiple CNN architectures has not been systematically evaluated for pasture applications, though multi-architecture approaches have shown promise in other agricultural domains. The relative merits of different ensemble integration strategies, from simple averaging to sophisticated stacking approaches with meta-learning, require empirical assessment in previous PB monitoring domains. Perhaps most importantly, the extent to which prediction accuracy varies across grass species, farms, and temporal scales remains poorly characterised, limiting confidence in model generalisation beyond training conditions.

The present study addresses these knowledge gaps through systematic empirical and operational evaluation of established ML approaches applied to a context where their combined performance has not previously been assessed: ground-level smartphone RGB imagery for PB prediction across commercial P-BDS. While TL and ensemble methods are individually well-established in agricultural computer vision, their systematic evaluation for ground-level pasture imagery spanning multiple farms, species, and seasons, and including explicit assessment of spatial generalisation, has not been reported in the literature. Three research questions guided this investigation: (1) Can transfer learning with stacking ensembles achieve operational accuracy for pasture biomass prediction from smartphone imagery across diverse dairy farms? (2) What is the relative contribution of deep learning features, hand-crafted features, and weather variables to prediction accuracy? (3) How does model performance vary across different grass species, seasonal conditions and farm environments? To address these questions, the primary objective was to identify optimal modelling configurations for automated PB estimation from smartphone imagery suitable for operational deployment in commercial P-BDS through comprehensive comparison of 21 ML configurations including baseline models, end-to-end DL, TL with four CNN architectures, multi-architecture ensembles, hyperparameter optimisation, stacking approaches, data augmentation strategies, and grass-specific modelling.

2. Materials and methods

2.1. Study area and data sources

The study was conducted across 15 commercial dairy farms in New South Wales, Australia, representing diverse climatic conditions, soil types, and management practices typical of temperate P-BDS. Farms spanned coastal, tableland, and inland regions with mean annual rainfall ranging from 650 to 1100 mm. The dataset encompassed measurements collected between April 2022 and July 2024, capturing substantial temporal and spatial variability in pasture production across multiple growing seasons and environmental conditions.

Pasture biomass measurements were obtained using the RPM, with 70 to 100 individual point measurements taken per paddock per assessment day along transect paths to capture within-paddock spatial heterogeneity. Compressed sward heights (CSH) were recorded and converted to biomass using farm-specific calibration equations. These calibration equations were established through destructive sampling and oven-drying procedures at each farm, providing ground truth PB estimates with established accuracy for Australian dairy pasture systems [24,45]. These point measurements were averaged to provide a single representative PB value for each paddock on each assessment date. The final dataset comprised 3291 unique paddock assessments, with values ranging from 1066 to 4622 kg DM ha^{-1} (mean = 2740 kg DM ha^{-1}, standard deviation = 638 kg DM ha^{-1}, coefficient of variation = 23.3%).

Paddock images were captured using standardised smartphone photography protocols to ensure consistency across farms and dates. One representative photograph was taken per paddock on each measurement occasion, acquired at ground level from a consistent height and viewing angle. The image location within each paddock was operator-selected on each visit to represent the general visual condition of the sward, consistent with standard commercial dairy farm monitoring practice, though this introduces potential for operator-dependent subjectivity in image placement. Paddocks across the 15 farms ranged from 1.04 to 12.11 ha (mean 3.92 ha, median 3.69 ha), and within-paddock spatial heterogeneity in biomass was partially mitigated by the RPM transect protocol averaging 70 to 100 point measurements per paddock. Images were stored in JPEG or HEIC format and subsequently converted to JPEG for processing compatibility. Each image was captured on the same day as RPM measurements to ensure temporal correspondence between visual appearance and biomass ground truth. The temporal distribution of measurements aligned with dominant pasture species, with ryegrass assessments conducted from May to November accounting for 2058 observations (62.6%), and kikuyu assessments from December to April comprising 1231 observations (37.4%).

Daily weather data were obtained for each farm location from the nearest weather station. Eight weather variables were extracted for the image capture date: maximum temperature (T.Max, °C), minimum temperature (T.Min, °C), rainfall (Rain, mm), pan evaporation (Evap, mm), solar radiation (Radn, MJ m^{-2}), vapour pressure (VP, hPa), relative humidity at maximum temperature (RHmaxT, %), and relative humidity at minimum temperature (RHminT, %). These variables provided environmental information for each image, potentially explaining visual ambiguity arising from recent precipitation, temperature stress, or growth conditions not directly visible in single time-point photographs. Missing weather data, which occurred for 27 observations (0.82% of the dataset), were imputed using mean substitution. These variables were selected to capture key environmental drivers of pasture growth and canopy spectral properties that could provide valuable contextual information for biomass prediction. All analyses were conducted in Python (version 3.11.4) using scikit-learn (version 1.3.0), TensorFlow (version 2.13.0), and XGBoost (version 1.7.6) libraries.

2.2. Image pre-processing

All images underwent standardised pre-processing prior to feature extraction to isolate pasture-relevant visual information and ensure compatibility with CNN architectures. Pre-processing comprised two sequential operations applied consistently across the entire dataset. Images originally in HEIC format were converted to JPEG using standard conversion utilities. An automatic semantic segmentation approach using the SegFormer-B0 model pre-trained on the ADE20K dataset was implemented to identify and remove sky pixels (class 2) and tree pixels (class 4) from paddock images [46,47]. SegFormer-B0 was selected over alternative segmentation architectures (e.g., DeepLab, U-Net) due to its computational efficiency, requiring minimal processing time (~0.5 s per image on standard hardware), whilst maintaining high segmentation accuracy (mIoU = 0.37 on ADE20K validation set). The transformer-based architecture of the model provided robust performance without extensive hyperparameter tuning, facilitating reproducible implementation across the full dataset. Segmentation predictions were generated with default confidence thresholds, with all pixels classified with probability >0.5 assigned to their respective classes [46,47].

Identified sky and tree regions were removed through cropping to the bounding box of remaining vegetation content rather than masking to zero. This cropping approach was adopted to maximise the proportion of informative pixels fed to CNN architectures whilst eliminating background distractors. Although cropping to bounding boxes removes some spatial context, this trade-off was considered acceptable because PB estimation relies primarily on vegetation texture and colour properties rather than absolute spatial positioning within the frame [31], pre-trained CNN features from ImageNet capture vegetation patterns independently of image dimensions [17,48], and preliminary testing indicated minimal performance difference between cropping and masking approaches, whilst cropping reduced computational overhead during subsequent feature extraction.

Following segmentation and cropping, all images were resized to 224 × 224 pixels using bilinear interpolation to standardise spatial dimensions and match the input requirements of pre-trained CNN architectures [49]. Pixel intensity values were normalised to the range 0 to 1 by dividing 8-bit integer values by 255, ensuring compatibility with ImageNet pre-trained networks. No explicit illumination normalisation or colour constancy correction was applied during pre-processing. This decision reflected a deliberate choice to preserve the natural spectral characteristics of pasture imagery, as vegetation indices and CNN features were hypothesised to learn robust representations across varying lighting conditions present in the training data. The dataset encompassed images captured across multiple seasons, times of day, and weather conditions (Section 2.1), providing natural exposure to illumination variability. However, this approach introduces a potential limitation, as HSV colour features and vegetation indices may be sensitive to extreme illumination variations (e.g., early morning shadows, overcast conditions). Vegetation indices based on visible spectrum ratios, whilst designed to minimise atmospheric and soil background effects, can exhibit reduced sensitivity under highly variable lighting conditions [50]. Similarly, HSV colour space separates chromaticity from intensity but does not fully compensate for dramatic illumination changes without explicit normalisation [51]. The impact of illumination variability on model performance is addressed in Section 4.6 (Limitations and Future Research Directions), where illumination normalisation is identified as a priority for future refinement.

Additionally, no additional colour space transformations, histogram equalisation, or augmentation operations were applied during pre-processing to preserve the original spectral characteristics of pasture imagery. The pre-processing pipeline was designed for computational efficiency whilst maintaining image quality sufficient for subsequent feature extraction and model training.

2.3. Feature extraction

Four distinct feature types were extracted from each pre-processed image to provide complementary information for pasture biomass prediction: hand-crafted features, deep learning features from multiple CNN architectures, temporal metadata, and weather variables. The complete feature extraction pipeline generated 4959 features per image. To address high dimensionality relative to sample size (3291 images), mutual information regression was subsequently employed to select the most informative features for model training.

2.3.1. Hand-crafted image features

A set of 67 hand-crafted features was computed using established computer vision techniques implemented through OpenCV (version 4.5) and scikit-image libraries. These features were selected to capture spectral, textural, structural, and spatial properties known to correlate with vegetation biomass in grassland systems. Colour features comprised 30 dimensions quantifying spectral properties across red-green-blue (RGB) and hue-saturation-value (HSV) colour spaces. For RGB channels, seven statistical metrics were computed per channel (mean, standard deviation, median, 25th percentile, 75th percentile, skewness, kurtosis), yielding 21 RGB features. For HSV channels, three statistical metrics were computed per channel (mean, standard deviation, median), yielding 9 HSV features. These colour descriptors provide basic spectral characterisation of vegetation density and greenness distribution.

Vegetation indices specifically designed for green biomass

quantification from visible spectrum imagery were selected based on established performance in grassland monitoring studies. The Visible Atmospherically Resistant Index (VARI), Green Leaf Index (GLI), Normalised Green-Red Difference Index (NGRDI), and Excess Green Index (ExG) have demonstrated robust relationships with chlorophyll content and leaf area across diverse pasture conditions. These indices leverage spectral band ratios to minimise sensitivity to illumination variation and soil background effects. Summary statistics (mean, standard deviation, and quartiles for ExG) were computed for each vegetation index, yielding 11 vegetation index features including green ratio. Alternative indices such as Red-Green-Blue Vegetation Index (RGBVI) and Colour Index of Vegetation Extraction (CIVE) were not included as preliminary testing indicated redundancy with the selected indices, contributing minimal additional predictive information whilst increasing computational overhead.

Texture features totalling 14 dimensions were derived using two complementary approaches to capture spatial patterns related to canopy density and structure. Grey Level Co-occurrence Matrix (GLCM) analysis following Haralick et al. [52] quantifies second-order statistical texture properties by examining spatial relationships between pixel intensities. GLCM matrices were computed at three spatial distances (1, 3, and 5 pixels) and four orientations (0°, 45°, 90°, 135°) from greyscale-converted images. Five Haralick texture properties were extracted from each GLCM: contrast, dissimilarity, homogeneity, energy, and correlation, with mean and standard deviation computed across all distance-orientation combinations to provide rotation-invariant texture descriptors (10 features).

Additional texture characterisation employed Local Binary Pattern (LBP) analysis with 24 sampling points at radius 3 pixels using the uniform pattern variant [53]. LBP captures local microstructure patterns complementary to GLCM by encoding relationships between central pixels and their circular neighbourhoods, providing sensitivity to fine-scale texture variations associated with leaf density. Four summary statistics were computed from the LBP histogram: mean, standard deviation, maximum bin value, and low-frequency content (sum of first five histogram bins) (4 features). Alternative texture descriptors such as Gabor filters were not employed due to their high computational cost and requirement for extensive parameter tuning (filter orientations, frequencies, bandwidths) that would complicate reproducibility without demonstrated performance advantages for pasture imagery in preliminary testing.

Edge detection captures boundaries and discontinuities potentially related to pasture height transitions and density gradients. Edge statistics (mean and standard deviation) were computed from Canny edge detection with thresholds of 50 and 150. These threshold values were selected based on established computer vision practice for natural scene imagery, where lower threshold (50) captures weak edges whilst upper threshold (150) retains only strong edges, with their ratio of approximately 1:3 following the recommended guidelines established by Canny. Whilst these fixed thresholds may exhibit reduced sensitivity under extreme lighting conditions (harsh shadows, overcast skies), the inclusion of weather variables and temporal metadata in the feature set provides additional context to compensate for illumination variability. Sobel gradient magnitudes in horizontal and vertical directions were calculated, with overall gradient magnitude computed as the Euclidean norm of these components. Summary statistics (mean, standard deviation, 90th percentile) of gradient magnitude captured structural patterns potentially related to pasture density and height. The 90th percentile specifically captures strong gradient responses associated with pronounced canopy structure whilst reducing sensitivity to noise.

Spatial features totalling 5 dimensions assessed within-image heterogeneity through quadrant-based analysis. Each image was divided into four equal quadrants, with mean green channel intensity computed for each quadrant. The standard deviation of these four quadrant means quantified spatial variability in green biomass distribution. This coarse spatial measure provides a computationally efficient indicator of biomass patchiness. More granular spatial statistics such as variogram analysis or entropy-based measures could potentially capture finer-scale heterogeneity but would substantially increase computational complexity and feature dimensionality. Given the subsequent feature selection step and the availability of hierarchical spatial representations from CNN features, this simplified spatial descriptor was considered sufficient for initial feature extraction. All hand-crafted features were standardised to zero mean and unit variance using StandardScaler prior to model training to ensure comparable scaling across features with different measurement units and ranges.

2.3.2. Deep learning features

Transfer learning was implemented using four pre-trained CNN architectures from the Keras Applications library: Residual Network 50 (ResNet50) [17], Visual Geometry Group 16 (VGG16) [54], MobileNetV2 [55], and DenseNet121 [56].

The combination of four architectures was hypothesised to capture complementary visual patterns relevant for biomass estimation, with ResNet50 and DenseNet121 capturing hierarchical features at multiple scales, VGG16 preserving fine-grained texture information, and MobileNetV2 providing efficient intermediate representations. This multi-architecture approach parallels ensemble techniques in traditional machine learning, where combining predictions from diverse models typically outperforms any single model by reducing variance.

All networks were initialised with ImageNet weights and employed with their classification heads removed, using global average pooling to generate fixed-length feature vectors. All convolutional layers remained frozen during feature extraction to preserve learned representations from ImageNet pre-training without task-specific fine-tuning. ImageNet pre-training provides general-purpose visual features learned from 1.2 million natural images, establishing hierarchical representations of edges, textures, and structural patterns. Whilst ImageNet contains primarily natural scenes rather than agricultural imagery, low-level and mid-level features such as edge detectors and texture filters transfer effectively to pasture monitoring tasks. However, high-level semantic features optimised for object recognition may not fully capture pasture-specific patterns related to biomass density and canopy architecture, representing a potential limitation of the transfer learning approach.

ResNet50 comprises 50 convolutional layers organised into residual blocks with skip connections that facilitate training of very deep networks by mitigating vanishing gradient problems. The architecture was loaded with ImageNet pre-trained weights, with final fully connected classification layers removed. Global average pooling was applied to the final convolutional layer output to produce a 2048-dimensional feature vector per image. VGG16 implements a simpler sequential architecture comprising 16 layers with small 3×3 convolutional filters and max pooling operations. The network emphasises local texture patterns through multiple stacked convolutions with small receptive fields. Following the same protocol as ResNet50, pre-trained weights were loaded, classification layers were removed, and global average pooling was applied to the final convolutional output to generate 512-dimensional feature vectors. MobileNetV2 employs inverted residual structures with linear bottlenecks designed for computational efficiency on mobile devices, producing 1280-dimensional feature vectors through global average pooling. DenseNet121 implements dense connectivity patterns where each layer receives feature maps from all preceding layers, facilitating feature reuse and gradient flow, yielding 1024-dimensional feature vectors. The dense connectivity pattern of DenseNet121 is particularly suited to grass canopy texture analysis, as feature reuse across network depth enables simultaneous capture of fine-grained leaf-level microstructure and coarser canopy-level density patterns relevant to biomass estimation.

The features were extracted in batch mode with batch size 32 to balance computational efficiency and memory constraints and extracted DL features were standardised using StandardScaler before concatenation with other feature types, yielding a total of 4864 DL features (2048

+ 512 + 1280 + 1024).

2.3.3. Temporal metadata and weather features

Temporal data was encoded through 20 metadata features derived from image capture dates. Season was represented using one-hot encoding across four categories (Summer: December to February, Autumn: March to May, Winter: June to August, Spring: September to November). Continuous temporal information was provided through day-of-year scaled to the range 0 to 1, enabling models to learn cyclic seasonal patterns within years. These metadata features enabled models to account for systematic differences in pasture growth patterns across farms and seasons that may not be fully captured through visual appearance alone. The eight weather variables measured on the image capture date for each farm location were included as exogenous predictors to provide environmental context for PB estimation. Weather variables were standardised (z-score normalisation) to zero mean and unit variance using StandardScaler to ensure comparable scaling across variables with different measurement units and ranges. The complete feature set comprised 4959 features: 67 hand-crafted features, 4864 DL features, 20 metadata features, and 8 weather features, with 4951 features available when weather was excluded.

2.4. Feature selection

Mutual information regression was employed for feature selection, ranking all features according to their mutual information with the target PB values [57]. Mutual information quantifies the reduction in uncertainty about the target variable provided by knowledge of each feature, capturing both linear and non-linear dependencies without assuming specific functional forms. We implemented this using SelectKBest with mutual information estimated via k-nearest neighbours density estimation (k = 3). Two feature selection thresholds were applied based on dataset size: 120 features for the full dataset (general and weather-integrated models), and 80 features for grass-specific models with smaller sample sizes. For the primary model with weather integration, retaining 120 features ensured a ratio of approximately 22 training samples per feature (2632 training samples divided by 120 features), which aligns with general guidelines for avoiding overfitting in tree-based ensemble methods whilst preserving sufficient dimensionality to capture complex visual and contextual patterns. To assess sensitivity to this threshold, the best-performing stacking ensemble model was additionally evaluated using 60 and 240 features.

2.5. Experimental design

Twenty-one distinct modelling approaches were systematically evaluated across seven experimental phases to enable direct comparison of feature types, model architectures, ensemble strategies, and data augmentation approaches. All experiments employed a random 80:20 train-test split with random state 42 for reproducibility. The training set comprised 2632 images whilst the test set contained 659 images held out exclusively for final performance evaluation. Table 1 confirms that this split produced well-balanced biomass distributions across both partitions, with no significant difference in mean values (two-sample *t*-test t = −1.468, p = 0.142). Cross-validation employed 5-fold stratified sampling within the training set for hyperparameter tuning and meta-learner training. The experimental configurations are summarised in Appendix Table A1, which details model specifications, feature types, hyperparameters, and experimental objectives for each of the 21 approaches evaluated.

Table 1
Comparison of pasture biomass distributions in training and test sets.

Characteristic	Training Set (n = 2632)	Test Set (n = 659)
Mean (kg DM ha^{-1})	2732	2772
Standard Deviation (kg DM ha^{-1})	640	630
Minimum (kg DM ha^{-1})	1066	1141
Maximum (kg DM ha^{-1})	4622	4405
Median (kg DM ha^{-1})	2698	2731
25th Percentile (kg DM ha^{-1})	2278	2321
75th Percentile (kg DM ha^{-1})	3164	3226

Two-sample *t*-test: t = −1.468, p = 0.142 (no significant difference.

2.5.1. Baseline models (experiments 1-3)

Baseline performance was established using only the 67 hand-crafted image features without deep learning integration (Table A1). Three regression algorithms were evaluated with initial hyperparameters [58–60]. These baseline experiments established the performance ceiling achievable without CNN feature extraction, providing a reference point for evaluating the contribution of transfer learning approaches.

2.5.2. Deep learning from scratch (experiments 4-7)

Four DL architectures were trained from scratch on the dataset to evaluate whether custom networks could learn effective representations without TL (Table A1). Experiment 4 implemented end-to-end fine-tuning of ResNet50, initialising the network with ImageNet weights but allowing all layers to be trainable. Two additional dense layers (256 units with ReLU activation and 50% dropout, followed by 128 units with ReLU activation and 30% dropout) were added before the final single-unit output layer. The model was trained for 50 epochs with batch size 16, Adam optimiser with learning rate 0.0001, and mean squared error loss.

Experiment 5 implemented a custom attention CNN incorporating spatial and channel attention mechanisms to learn where to focus within images [61,62]. The network comprised four convolutional blocks with residual connections, followed by spatial attention modules that learn location-specific feature weightings and channel attention modules that emphasise informative feature channels. Experiment 6 implemented a density estimation network attempting pixel-level biomass prediction at 56 × 56 resolution using fully convolutional architecture, though no ground truth spatial labels were available for supervision. Experiment 7 applied Vision Transformer (ViT) architecture [63], dividing images into 16 × 16 pixel patches (196 patches total) with each patch linearly projected to embedding space and processed through 8 transformer encoder layers with multi-head attention.

All architectures trained from scratch employed early stopping based on validation loss (patience = 10 epochs). Whilst the training dataset size was expected to be insufficient for these architectures, which typically require tens of thousands of samples for effective learning without pre-training, these experiments were included to provide empirical benchmarks quantifying the performance gap between end-to-end learning and transfer learning approaches.

2.5.3. Transfer learning, Multi-CNN ensembles and hyperparameter optimisation (Experiment 8-14)

Experiments 8 through 11 evaluated individual CNN architectures for TL-based feature extraction (Table A1). Pre-trained CNN features were extracted with frozen convolutional layers, and RF regressors were trained on these features. These experiments enabled direct comparison of architectural choices for TL, quantifying the relative effectiveness of different CNN techniques for pasture imagery feature extraction. Experiment 12 integrated features from all four CNN architectures (4864 dimensions) with hand-crafted features (67 dimensions) and temporal metadata (20 dimensions), yielding 4951 total features prior to mutual information selection of the top 120 features. Base model ensemble comprised RF, XGBoost, and GBM with initial hyperparameters, combined through simple averaging with equal weights (0.333 each). Experiment 13 incorporated the 8 weather features, expanding the feature space to 4959 dimensions prior to selection. The same multi-CNN ensemble architecture was employed, enabling assessment of weather data contribution through direct comparison with Experiment 12. Experiment 14 performed systematic hyperparameter

optimisation through RandomizedSearchCV with 50 iterations and 3-fold cross-validation on the multi-CNN feature set with weather integration. Search spaces are detailed in Table A1. Optimal configurations were selected based on cross-validation R^2 performance, with these tuned hyperparameters employed in all subsequent experiments.

2.5.4. Ensemble methods (experiments 15-17)

Three ensemble integration strategies were systematically compared using optimised base models from Experiment 14 (Table A1). Experiment 15 implemented weighted ensemble with optimised coefficients determined through constrained optimisation minimising MAE using Sequential Least Squares Programming (SLSQP) algorithm. Experiments 16 and 17 employed stacked generalisation with Ridge regression meta-learning. For the stacking process, three base models (RF, XGBoost, GBM) with tuned hyperparameters were trained using 5-fold cross-validation to generate out-of-fold predictions. For each fold, models were trained on four folds and predictions generated for the held-out fold, producing out-of-fold predictions for all training samples. These out-of-fold predictions formed a 2632×3 meta-feature matrix. A Ridge regression meta-learner (alpha = 1.0) was trained on this meta-feature matrix to learn optimal linear combination weights through ordinary least squares with L2 regularisation. For test set predictions, base models were retrained on the full training set and their predictions combined using learned meta-learner weights. This approach prevents information leakage by ensuring the meta-learner never sees predictions from base models trained on the same data used for those predictions. Experiment 16 excluded weather features (4951 features selected to 120), whilst Experiment 17 incorporated weather variables (4959 features selected to 120), enabling direct quantification of weather data contribution to prediction accuracy.

2.5.5. Data augmentation (Experiment 18)

Training data augmentation was evaluated to determine whether synthetically expanding the training set could improve model generalisation (Table A1). Augmentation transformations included rotations (±20°), horizontal and vertical flips (50% probability), width and height shifts (±10%), and brightness adjustments (0.8 to 1.2, ±20%). An augmentation factor of 3 was applied, generating two synthetic variants for each original training image, expanding the effective training set from 2632 to 7893 images (67% synthetic). Augmented images were processed through ResNet50 feature extraction, with resulting features used to train RF regression using tuned hyperparameters from Experiment 14. Test set evaluation used original unaugmented images to assess generalisation performance.

2.5.6. Grass-specific models (experiments 19-21)

Experiments 19 through 21 evaluated species-specific models to account for biological differences between cool-season ryegrass and warm-season kikuyu pastures. Images were partitioned temporally based on dominant pasture species. Experiment 19 trained a ryegrass-specific model on May to November data (2058 samples: 1646 training, 412 test) using the full multi-CNN methodology with weather integration and mutual information selection retaining 80 features (reduced k due to smaller sample size). Experiment 20 trained a kikuyu-specific model on December to April data (1231 samples: 984 training, 247 test) using identical methodology. Experiment 21 implemented a combined grass-specific ensemble that routed test samples to the appropriate species-specific model based on image capture month, enabling comparison of universal versus species-specific modelling approaches. Model architecture, hyperparameters, and feature selection procedures remained consistent with Experiment 17 to enable direct performance comparison.

2.6. Model evaluation and cross-validation

Model performance was evaluated using three complementary metrics: coefficient of determination (R^2), mean absolute error (MAE, kg DM ha^{-1}), and root mean squared error (RMSE, kg DM ha^{-1}). For stacked generalisation models, out-of-fold predictions from 5-fold cross-validation were used to train the Ridge meta-learner, ensuring meta-learner training data were independent of base model training data. Cross-validation consistency was assessed by calculating mean and standard deviation of performance metrics across the five folds, with low standard deviation indicating stable performance across different training-validation splits. Feature importance for tree-based models was quantified using mean decrease in impurity, ranking features according to their contribution to prediction accuracy.

Leave-one-farm-out cross-validation (LOFO-CV) was conducted on the best-performing model (Experiment 17) to assess generalisation capability to unseen farm environments. For each of the 15 farms, a model was trained on data from the remaining 14 farms and evaluated on the held-out farm. This procedure mimics the operational scenario where a model trained on multiple farms must predict PB at a new farm, providing rigorous assessment of spatial transferability. Performance was additionally disaggregated by season (Summer, Autumn, Winter, Spring), farms, and geographic region (North Coast, Mid Coast, South Coast) to characterise prediction accuracy across diverse conditions. All statistical analyses and model training were conducted using custom Python scripts implementing scikit-learn, TensorFlow/Keras, and XGBoost libraries. Fig. 1 summarises the complete methodological workflow from data collection through feature extraction, model training, and ensemble integration.

3. Results

The evaluation of 21 ML configurations across the 15 farms examined in this study revealed substantial performance variation, with R^2 ranging from 0.140 to 0.561 for models evaluated on identical train-test splits (2632 training samples, 659 test samples). Stacking ensemble methods integrating TL features from four CNN architectures with weather data achieved accuracy for PB prediction ($R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, RMSE = 437 kg DM ha^{-1}), representing a 48% improvement over baseline models using hand-crafted features only In contrast, deep learning approaches trained from scratch performed poorly as anticipated given the limited training data, with four of five such methods achieving negative R^2 values or near-zero performance. Grass-specific models (Experiments 19-21) trained on separate seasonal subsets (ryegrass: May-November; kikuyu: December-April) demonstrated species-level differences in prediction capacity but employed different train-test partitions and are reported separately. Leave-one-farm-out cross-validation of the best model revealed variable spatial generalisation across farm environments.

3.1. Comparative performance of modelling approaches

Twenty-one modelling configurations were evaluated across seven methodological phases: baseline models (Experiments 1-3), DL from scratch (Experiments 4-7), TL (Experiments 8-11), multi-CNN ensembles (Experiments 12-13), hyperparameter optimisation (Experiment 14), advanced ensemble methods (Experiments 15-17), data augmentation (Experiment 18), and grass-specific models (Experiments 19-21). Fifteen of the eighteen experiments using identical train-test splits are compared in Table 2, with performance among competitive models ranging from $R^2 = 0.140$ to $R^2 = 0.561$. Three deep learning from scratch approaches (Experiments 4, 5, and 7) that collapsed due to insufficient training data are excluded from Table 2 but included in Appendix Table A2 for completeness. Grass-specific models (Experiments 19-21) are excluded from Table 2 as they employed different seasonal train-test partitions.

The stacking ensemble with weather integration (Experiment 17) achieved the highest performance ($R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, RMSE = 437 kg DM ha^{-1}) as shown in Fig. 2, representing a 48% improvement in R^2 over the best baseline model using hand-crafted

Fig. 1. Methodological workflow for pasture biomass prediction using stacking ensemble with multi-source features (hand-crafted, deep learning, metadata, weather). The pipeline includes SegFormer-B0 preprocessing, parallel feature extraction from four CNN architectures (ResNet50, VGG16, MobileNetV2, DenseNet121), mutual information-based selection (4959 → 120 features), and stacking ensemble training with Ridge regression meta-learner combining three base models through 5-fold cross-validation on 3291 paddock assessments from 15 farms.

features only (Experiment 2: $R^2 = 0.380$). TL approaches consistently outperformed DL trained from scratch, with the multi-CNN ensemble (Experiment 12: $R^2 = 0.495$) exceeding all single-architecture TL models. Hyperparameter optimisation provided incremental gains (Experiment 14: $R^2 = 0.538$ vs Experiment 13: $R^2 = 0.521$), whilst stacking ensemble architecture yielded additional improvement through optimal linear combination of base learner predictions. DL models trained from randomly initialised weights failed systematically. Four of five such approaches achieved negative R^2 values or below-baseline performance. The custom attention CNN exhibited the weakest performance (Experiment 5, Table A2) and the Vision Transformer achieving near-zero predictive capacity (Experiment 7, Table A2) despite 3352 s training time. Data augmentation (Experiment 18) similarly failed due to feature dimension mismatch between augmented images and the original feature extraction pipeline. Grass-specific models employed separate seasonal train-test partitions. The ryegrass-specific model achieved $R^2 = 0.586$ on May-November samples, whilst the kikuyu-specific model attained $R^2 = 0.391$ on December-April samples. Experiment 21 combined both models but used fundamentally different evaluation protocols.

3.2. Best model architecture and performance

The optimal model (Experiment 17) employed a two-layer stacking ensemble architecture. Three base learners, RF, XGBoost, and GBM, were trained with optimised hyperparameters identified through RandomizedSearchCV (Experiment 14). Out-of-fold predictions generated via 5-fold cross-validation on the training set served as meta-features for

Table 2
Comparative performance of modelling approaches for pasture biomass prediction ranked by $R.^2$.

Rank	Experiment	Model	Category	R^2	RMSE (kg DM ha^{-1})	MAE (kg DM ha^{-1})	Notes
1	17	Stacking + Weather	Advanced Ensemble Methods	0.561	437	351	Best model
2	15	Weighted Ensemble	Advanced Ensemble Methods	0.549	442	355	Grid search weights
3	14	Hyperparameter Optimised Ensemble	Hyperparameter Optimisation	0.538	448	361	RandomizedSearchCV
4	16	Stacking No Weather	Advanced Ensemble Methods	0.530	451	364	Ridge meta-learner
5	13	Multi-CNN + Weather	Multi-CNN Ensembles	0.521	456	368	Simple averaging
6	12	Multi-CNN Ensemble	Multi-CNN Ensembles	0.495	468	378	No weather
7	11	DenseNet121 Transfer Learning	Transfer Learning	0.450	489	390	Single CNN
8	2	XGBoost Baseline	Baseline Models	0.380	519	410	Hand-crafted features only
9	3	Gradient Boosting Baseline	Baseline Models	0.369	523	412	Hand-crafted features only
10	1	Random Forest Baseline	Baseline Models	0.357	528	421	Hand-crafted features only
11	9	VGG16 Transfer Learning	Transfer Learning	0.347	532	426	Single CNN
12	10	MobileNetV2 Transfer Learning	Transfer Learning	0.331	539	431	Single CNN
13	6	Density Estimation Network	Deep Learning from Scratch	0.223	581	468	No spatial labels
14	8	ResNet50 Transfer Learning	Transfer Learning	0.152	606	495	Single CNN
15	18	Training Augmentation	Data Augmentation	0.140	611	498	Feature mismatch

Experiments 4 (End-to-End ResNet50, $R^2 = 0.010$), 5 (Custom Attention CNN, $R^2 = -0.781$), and 7 (Vision Transformer, $R^2 = 0.007$) and grass-specific models (Experiments 19-21) are excluded from this table; full results for all 21 experiments are provided in Appendix Table A2. All experiments used identical random 80:20 train-test splits (2632 training samples, 659 test samples, random state 42). Performance metrics computed on held-out test set. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error.

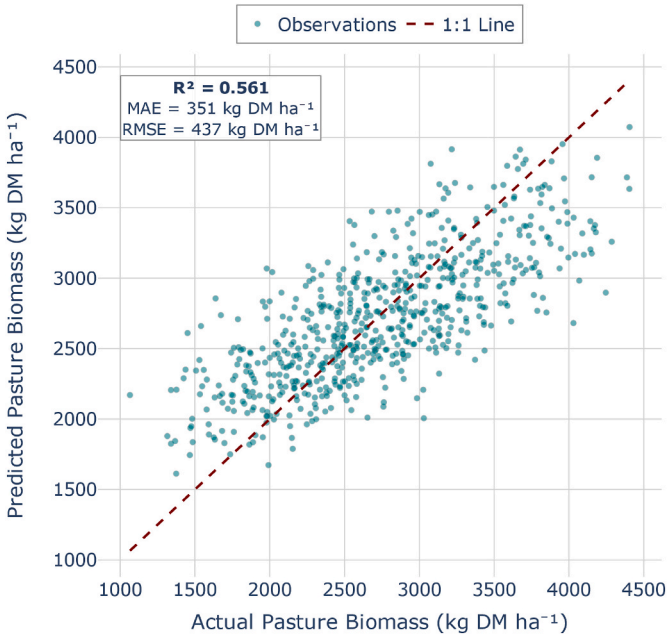


Fig. 2. Predicted versus actual pasture biomass for the best model (Experiment 17: stacking ensemble with weather integration) evaluated on the held-out test set (n = 659). The dashed line represents perfect prediction (1:1). R^2: coefficient of determination; MAE: mean absolute error; RMSE: root mean squared error.

a Ridge regression meta-learner ($\alpha = 1.0$), which learned optimal linear combination weights through L2-regularised ordinary least squares. Base learner performance during 5-fold cross-validation demonstrated consistent predictive capacity across folds. GBM achieved the highest mean cross-validation performance ($R^2 = 0.487 \pm 0.031$, MAE = 362 ± 9 kg DM ha^{-1}), followed by XGBoost ($R^2 = 0.484 \pm 0.036$, MAE = 365 ± 10 kg DM ha^{-1}) and RF ($R^2 = 0.420 \pm 0.034$, MAE = 386 ± 11 kg DM ha^{-1}). Low standard deviations across folds confirmed stable model training and robust generalisation capacity. The Ridge meta-learner assigned weights reflecting individual base learner performance: GBM received the highest weight (0.617), followed by XGBoost (0.340) and RF (0.155), with a learned intercept of −87.3 kg DM ha^{-1}. This weighted combination yielded test set performance ($R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, RMSE = 437 kg DM ha^{-1}) that exceeded all three base learners individually, confirming the value of ensemble integration for reducing prediction variance.

3.3. Feature importance analysis

Mutual information analysis identified the top 120 most informative features from the complete feature pool of 4959 dimensions for the best model. Hand-crafted vegetation indices and DL features from DenseNet121 dominated the importance rankings, with 10 hand-crafted features and 9 DL features comprising the top 20 positions (Fig. 3). Green ratio ranked highest overall (MI score = 0.133), followed by DenseNet121 Feature 427 (MI score = 0.132) and excess green index mean (MI score = 0.129). Among hand-crafted features, vegetation indices (green ratio, excess green index, GLI) and colour-based metrics (HSV saturation median, hue mean) exhibited stronger associations with PB than texture or edge features. DL features from DenseNet121 appeared eight times in the top 20, substantially exceeding contributions from ResNet50, VGG16, and MobileNetV2. Weather variables demonstrated moderate importance, with maximum temperature ranking 20th (MI score = 0.075), followed by evaporation (rank 21), solar radiation (rank 27), minimum temperature (rank 75), and vapour pressure (rank 80) within the top 120. Temporal metadata contributed modestly, with day-of-year ranking 32nd (MI score = 0.068). The balanced representation of hand-crafted and DL features in top rankings validated the ensemble integration approach, whilst the moderate importance of weather variables confirmed their complementary predictive value beyond image-derived features alone. Feature threshold sensitivity analysis on the best model yielded $R^2 = 0.558$, MAE = 352 kg DM ha^{-1}, RMSE = 438 kg DM ha^{-1} at 60 features; $R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, RMSE = 437 kg DM ha^{-1} at 120 features; and $R^2 = 0.568$, MAE = 350 kg DM ha^{-1}, RMSE = 433 kg DM ha^{-1} at 240 features, confirming that model performance was stable across this range and that the 120-feature threshold represented a computationally efficient choice without meaningful accuracy trade-off.

3.4. Temporal performance analysis

Model performance varied substantially across seasons, reflecting phenological differences in pasture growth patterns and image characteristics. The best model (Experiment 17) achieved highest accuracy during spring ($R^2 = 0.701$, MAE = 308 kg DM ha^{-1}), when pasture growth was most vigorous and vegetation indices exhibited strongest

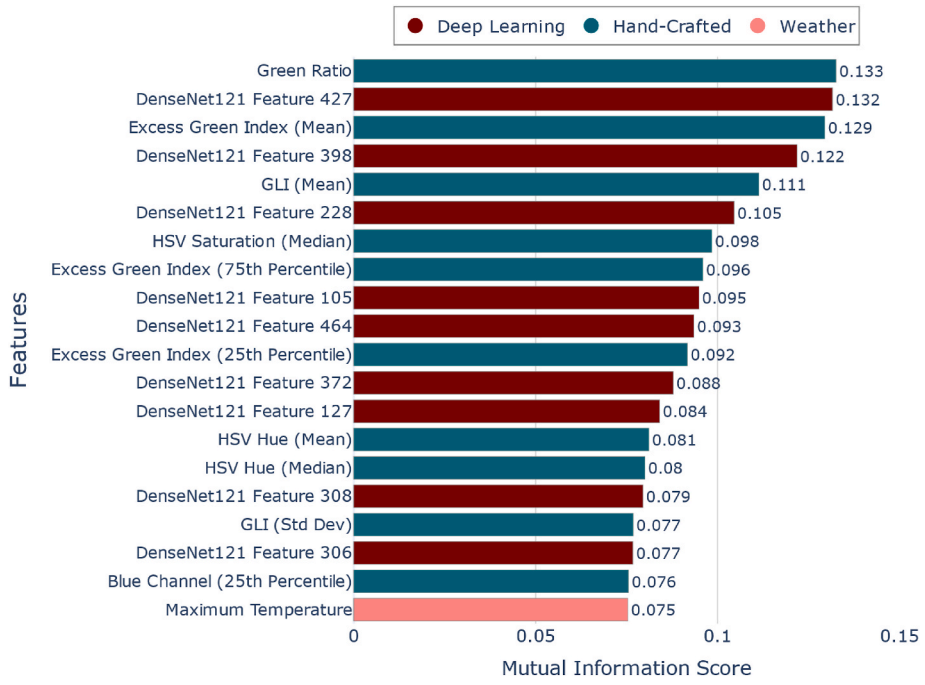


Fig. 3. Top 20 features ranked by mutual information score for pasture biomass prediction from the best model (Experiment 17). Features are colour-coded by type: deep learning features from pre-trained CNN architectures (maroon), hand-crafted image features including vegetation indices and colour metrics (teal), and weather variables (light coral). GLI: Green Leaf Index; HSV: Hue, Saturation, Value colour space. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

correlations with PB. Summer and autumn demonstrated similar moderate accuracy ($R^2 = 0.486$ and 0.484 respectively), with MAE values of 362 and 357 kg DM ha^{-1}, reflecting relatively stable prediction capacity across the active growth period.

Winter exhibited lowest performance ($R^2 = 0.456$, MAE = 375 kg DM ha^{-1}), consistent with slower growth rates, reduced green biomass, and challenging lighting conditions during shorter daylight periods. Despite seasonal variations in R^2, MAE remained below 380 kg DM ha^{-1} across all seasons, with RMSE ranging from 387 kg DM ha^{-1} in spring to 460 kg DM ha^{-1} in winter. The relatively consistent error margins throughout the annual production cycle (MAE range: 308-375 kg DM ha^{-1}) demonstrated operationally acceptable prediction accuracy, though the substantial improvement in spring R^2 (+0.245 compared to winter) indicated sensitivity to phenological conditions affecting vegetation spectral characteristics (Fig. 4).

3.5. Spatial performance and generalisation analysis

3.5.1. Test set performance by farm and region

Farm-level performance on the test set (n = 659) exhibited R^2 ranging from −0.806 to 0.577 across 13 farms (Table 3). Regional aggregation revealed consistent within-dataset performance: South Coast (n = 451, $R^2 = 0.564$, MAE = 331 kg DM ha^{-1}, RMSE = 414 kg DM ha^{-1}), North Coast (n = 109, $R^2 = 0.552$, MAE = 397 kg DM ha^{-1}, RMSE = 475 kg DM ha^{-1}), and Mid Coast (n = 99, $R^2 = 0.544$, MAE = 390 kg DM ha^{-1}, RMSE = 491 kg DM ha^{-1}). Regional performance differences were modest (R^2 range: 0.544-0.564, MAE range: 331-397 kg DM ha^{-1}), indicating relatively uniform prediction accuracy when test samples were drawn from farms represented in the training data.

3.5.2. Leave-one-farm-out cross-validation

LOFO-CV assessed spatial generalisation capacity by systematically withholding each farm and training on the remaining 14 farms (Table 4). Performance degraded substantially compared to within-dataset evaluation, with aggregate LOFO-CV achieving mean $R^2 = -0.009 \pm 0.781$, MAE = 399 ± 74 kg DM ha^{-1}, and RMSE = 488 ± 87 kg DM ha^{-1}. Individual farm R^2 ranged from −2.293 to 0.517, with five farms achieving $R^2 > 0.4$ but three exhibiting negative R^2 values indicating systematic prediction failure when these environments were excluded from training data.

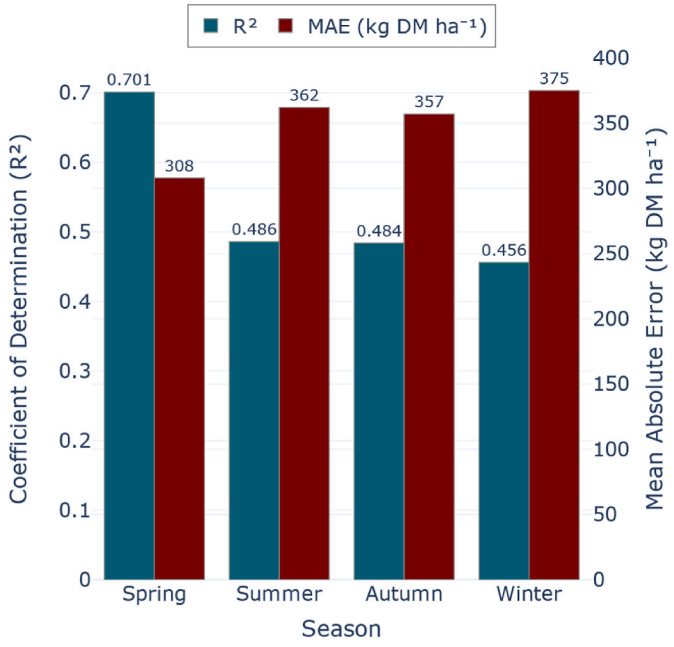


Fig. 4. Seasonal performance of the best model (Experiment 17) for pasture biomass prediction. Coefficient of determination (R^2, teal bars, left axis) and mean absolute error (MAE, maroon bars, right axis) are shown for each season. The dual y-axis format enables direct comparison of model accuracy and prediction error across the annual production cycle.

Regional LOFO-CV patterns differed markedly from test set performance. South Coast maintained moderate generalisation (mean R^2 =

Table 3
Test set performance by farm for the best model (Experiment 17).

Farm	Region	n	R^2	MAE (kg DM ha^{-1})	RMSE (kg DM ha^{-1})
Farm 6	South Coast	80	0.577	375	458
Farm 7	North Coast	27	0.558	435	520
Farm 1	South Coast	111	0.558	317	399
Farm 5	Mid Coast	43	0.520	330	423
Farm 8	South Coast	92	0.516	286	369
Farm 14	Mid Coast	47	0.512	460	567
Farm 15	North Coast	23	0.512	357	445
Farm 3	North Coast	35	0.499	445	509
Farm 13	South Coast	84	0.475	364	441
Farm 12	South Coast	84	0.367	325	406
Farm 2	Mid Coast	3	0.179	281	321
Farm 4	North Coast	24	−0.202	321	388
Farm 11	Mid Coast	4	−0.806	328	386

Farms sorted by R^2 (descending). Test set drawn from random 20% split of complete dataset (n = 659 from 3291 total samples). Farm numbering corresponds to order in sorted performance rankings.

Table 4
Leave-one-farm-out cross-validation results for the best model (Experiment 17).

Farm	Region	n	R^2	MAE (kg DM ha^{-1})	RMSE (kg DM ha^{-1})
Farm 3	North Coast	147	0.517	436	511
Farm 8	South Coast	445	0.450	337	425
Farm 6	South Coast	444	0.428	409	498
Farm 13	South Coast	469	0.397	376	472
Farm 9	Mid Coast	5	0.394	259	339
Farm 2	Mid Coast	10	0.347	282	332
Farm 1	South Coast	448	0.345	383	473
Farm 15	North Coast	132	0.339	375	487
Farm 7	North Coast	131	0.271	447	545
Farm 14	Mid Coast	236	0.225	560	673
Farm 5	Mid Coast	211	0.098	411	516
Farm 12	South Coast	452	0.013	421	511
Farm 4	North Coast	141	−0.362	487	603
Farm 11	Mid Coast	10	−1.303	410	499
Farm 10	Mid Coast	10	−2.293	394	441

Farms sorted by R^2 (descending). Each LOFO-CV fold withheld one farm as the test set whilst training on the remaining 14 farms (n = 3291 total samples across 15 farms). Three farms with n ≤ 10 may exhibit inflated performance variability due to limited sample sizes.

0.327, MAE = 385 kg DM ha^{-1}, RMSE = 476 kg DM ha^{-1}), North Coast showed reduced capacity (mean R^2 = 0.191, MAE = 436 kg DM ha^{-1}, RMSE = 536 kg DM ha^{-1}), whilst Mid Coast demonstrated poor generalisation (mean R^2 = −0.422, MAE = 386 kg DM ha^{-1}, RMSE = 467 kg DM ha^{-1}). The contrast between test set performance (R^2 = 0.561) and LOFO-CV (R^2 = −0.009) revealed substantial spatial dependency in model predictions. Whilst within-dataset performance remained consistent across regions (R^2 = 0.544-0.564), LOFO-CV exposed farm-specific characteristics that limited generalisation to completely novel environments. However, MAE remained operationally acceptable across both evaluations (test set: 331-397 kg DM ha^{-1}; LOFO-CV: 385-436 kg DM ha^{-1}), suggesting that prediction magnitude errors were relatively stable even when R^2 indicated poor linear correlation.

3.6. Grass-type-specific models

Three exploratory experiments evaluated grass-type-specific modelling using seasonal train-test partitions distinct from the random 80:20 split used in Experiments 1-18 and are therefore not directly comparable with results in Table 2. The ryegrass-specific model (Experiment 19, May-November, n = 2060) achieved R^2 = 0.586, MAE = 315 kg DM ha^{-1}, and RMSE = 398 kg DM ha^{-1}. The kikuyu-specific model (Experiment 20, December-April, n = 1231) achieved R^2 = 0.391, MAE = 369 kg DM ha^{-1}, and RMSE = 467 kg DM ha^{-1}. A combined ensemble (Experiment 21) that routed predictions through the appropriate grass-specific model based on season achieved R^2 = 0.914, MAE = 103 kg DM ha^{-1}, and RMSE = 193 kg DM ha^{-1} on a test set comprising 423 ryegrass and 236 kikuyu samples.

4. Discussion

This study evaluated 21 ML configurations for predicting PB from smartphone imagery across 15 commercial pasture-based dairy farms, revealing that TL with frozen pre-trained CNN features dramatically outperformed DL trained from scratch, multi-architecture feature integration captured complementary information, and stacking ensemble methods achieved optimal prediction accuracy through learned combination weights. However, substantial performance degradation in LOFO-CV exposed farm-specific dependencies limiting zero-shot generalisation to unseen farm environments. These findings establish both methodological best practices for agricultural computer vision and practical constraints for deploying smartphone-based PB monitoring systems in P-BDS.

4.1. Transfer learning superiority and feature extraction

The poor performance of DL architectures trained from scratch (Table 1, Experiments 4-7: R^2 = −0.781 to 0.223) contrasted sharply with frozen TL success (Experiments 8-11: R^2 = 0.152 to 0.450) shows that ImageNet-learned visual hierarchies provide relevant visual features for pasture canopy images despite domain shift from natural scenes to pasture canopies. Our training dataset of 2632 images proved fundamentally insufficient for learning the 23.5 million parameters in ResNet50 without catastrophic overfitting [64], whilst frozen convolutional layers leveraged visual concepts from 1.2 million ImageNet images to extract edges, textures, and structural patterns relevant to canopy density estimation [48,65]. This outcome aligns with established computer vision principles showing TL effectiveness when target domains share low-level visual primitives with source domains despite differing high-level semantic content [13,14].

Multi-architecture integration combining DenseNet121, ResNet50, VGG16, and MobileNetV2 features substantially exceeded single-architecture approaches, affirming that different CNN architectures capture complementary visual patterns relevant to PB estimation [27]. The dominance of DenseNet121 in feature importance rankings (Fig. 3) reflects its dense connectivity architecture enabling efficient feature reuse across network depth, allowing the model to learn hierarchical representations from coarse vegetation patterns to fine-grained canopy texture with fewer parameters than traditional sequential architectures [56]. Hand-crafted vegetation indices (green ratio, excess green index, GLI) provided comparable importance to DL features, occupying 10 of

the top 20 positions revealing that domain-specific colour transformations encode biologically meaningful information and phenomena such as chlorophyll content not fully captured by general-purpose CNN filters [20]. This balanced contributions from DL and hand-crafted features suggests hybrid feature engineering strategies may optimise performance when training data are limited, contrasting with pure end-to-end DL approaches that dominate large-scale vision tasks [17,54].

4.2. Ensemble learning and weather data integration

Stacking ensemble architecture achieved superior performance by learning optimal data-driven combination weights, demonstrating that meta-learning approaches exploit complementary strengths across diverse base learners. The weight distribution of the Ridge meta-learner (GBM: 0.617, XGBoost: 0.340, RF: 0.155) reflects the predictive capacity of each algorithm, with gradient boosting receiving substantially higher weights, indicating that sequential boosting captures pasture-biomass relationships more effectively. Weather integration contributed meaningful but modest improvement (5.9% gain), with maximum temperature as the most informative meteorological variable, confirming that climatic drivers modulate growth rates beyond visual canopy patterns [43]. However, the limited magnitude of weather contribution suggests image-derived features already capture substantial environmental information through phenological effects on canopy structure.

Comparing our stacking ensemble performance ($R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, RMSE = 437 kg DM ha^{-1}) with published smartphone-based approaches reveals competitive accuracy. Schaefer and Lamb [31] reported $R^2 = 0.52$ using support vector machines with hand-crafted texture features, whilst Näsi et al. [30] achieved $R^2 = 0.48$ with RF regression on smartphone spectral indices. Our integration of TL, multi-architecture features, and ensemble methods demonstrates meaningful improvement over these established baselines whilst maintaining smartphone accessibility. More complex systems combining proximal sensing with satellite data achieve higher accuracy ([22]: $R^2 = 0.86$ [23];: $R^2 = 0.87$), but require substantially greater technical infrastructure and data processing pipelines. Our RMSE of 437 kg DM ha^{-1} compares favourably with satellite-based remote sensing studies reporting RMSE values between 226 and 347 kg DM ha^{-1} for different pasture species [66] and falls well below multi-site RPM validation showing RMSE of approximately 904 kg DM ha^{-1} under diverse field conditions [67].

4.3. Temporal dynamics and seasonal performance

Substantial seasonal variation in model performance (Section 3.4, Fig. 4) reflects the complex interplay between pasture phenology, species composition, and environmental conditions affecting visual-biomass relationships. Spring superiority could have stemmed from vigorous cool-season ryegrass growth creating uniform vertical canopy architecture optimal for smartphone-based estimation. During peak spring growth, strong correlations between vegetation indices and standing biomass emerge as temperate species exhibit rapid leaf area expansion with consistent structural characteristics, whilst favourable solar radiation and longer photoperiods ensure high-quality image capture, collectively facilitating robust feature extraction.

Winter performance degradation likely results from multiple confounding factors. Species transition periods create heterogeneous mixed-species canopies as ryegrass dormancy coincides with emerging kikuyu growth, disrupting uniform visual patterns. Reduced solar radiation and shorter photoperiods compromise image quality through inconsistent lighting, whilst slower growth rates weaken temporal coupling between visual appearance and biomass accumulation. Summer and autumn exhibited similar intermediate performance, suggesting stable prediction during active growing seasons.

The grass-specific ensemble (Experiment 21) demonstrated proof-of-concept potential for species-tailored modelling, with substantial performance improvement confirming that ryegrass and kikuyu exhibit fundamentally different visual-biomass relationships [68]. Ryegrass uniformity through upright tillering could have enabled stronger predictability, contrasting with kikuyu prostrate stoloniferous growth creating horizontal canopy architecture that obscures direct height-biomass relationships [25,69]. This species-level variation aligns with previous findings showing marked performance reductions when species information was removed, with explained variance falling from 73% to 32% [70]. Taugourdeau et al. [71] showed that low cost cameras could explain about 60% of biomass variability, but their approach lacked the species specific calibration necessary for robust deployment on commercial farms. The present study addresses this limitation by explicitly modelling species level variation and demonstrating its measurable impact on predictive accuracy. However, practical application requires either a priori knowledge of dominant grass species or automated species identification capabilities, suggesting that general-purpose models accepting mixed-species canopies may prove more feasible for widespread adoption despite moderate accuracy trade-offs.

4.4. Spatial generalisation and model limitations

LOFO-CV results (Table 4, Section 3.5.2) exposed severe spatial generalisation constraints, with performance collapsing when predicting unseen farms despite strong within-dataset accuracy. This degradation can be attributed to three distinct and interacting sources.

First, data distribution shift arising from farm-specific management practices created distinct visual environments that models could not transfer across. Varied fertiliser rates, grazing rotations, and stocking densities produced sward conditions with different greenness, density, and structural characteristics that were farm-specific rather than universally predictive of PB. Second, biological differences in species composition and phenology introduced systematic variation in visual-PB relationships across farms and regions. Varying ryegrass cultivars and kikuyu establishment ratios created fundamentally different canopy architectures, whilst phenological timing differences across the 600 km coastal gradient meant that the same visual appearance corresponded to different PB values depending on local growth stage.

Third, sensor and viewpoint limitations inherent to 2D smartphone RGB imagery constrained cross-regional transferability. RGB sensors cannot directly capture canopy height, a primary PB determinant [72], and instead rely on colour and texture as proxies whose relationship with PB is modulated by illumination, camera angle, and local soil colour. Subtle between-farm variation in image capture protocols introduced additional systematic biases that further constrained transferability.

Beyond these three sources, a fundamental question is whether an intrinsic information bottleneck constrains achievable accuracy regardless of model complexity. RPM calibration relationships themselves typically explain only 70-80% of PB variance [45], establishing a theoretical ceiling on prediction accuracy. Furthermore, RGB imagery encodes only a 2D colour projection of a 3D canopy structure [73], meaning that information required to reliably predict PB across diverse farm environments may be fundamentally absent from the input signal. Increasing model complexity without addressing this sensor-level limitation is unlikely to substantially improve cross-regional generalisation.

The substantial LOFO-CV degradation represented a critical constraint for practical application, as farms would require models generalising beyond training locations without site-specific calibration. However, relatively stable MAE values suggested prediction magnitude errors remained acceptable for management decisions despite poor correlation, indicating systematic offset biases rather than complete failure. Potential mitigation strategies could include transfer learning fine-tuning with modest farm-specific calibration data, continual learning updating ensemble weights as measurements accumulate, or

active learning targeting challenging conditions [24].

4.5. Methodological novelty and practical implications

This study made three primary methodological contributions. The evaluation of 21 configurations spanning baseline ML, DL from scratch, TL with multiple architectures, ensemble methods, and grass-specific approaches provided comprehensive empirical comparison rarely documented in agricultural remote sensing, establishing that frozen TL with multi-architecture integration and stacking ensemble methods represented current best practice for smartphone-based PB estimation with moderate training data. Explicit demonstration that frozen features outperformed trainable architectures challenged assumptions favouring end-to-end learning, revealing data-efficiency advantages critical for agriculture where ground-truth collection remains labour-intensive. Rigorous LOFO-CV evaluation exposing severe generalisation constraints provided essential context often absent from ML studies reporting only within-dataset performance, potentially overestimating practical utility. The integration of hand-crafted vegetation indices, DL features from multiple architectures, temporal metadata, and weather variables captured complementary aspects of pasture conditions not achievable through visual features alone. Feature importance analysis (Fig. 3) confirmed this multimodal approach, with vegetation indices and DenseNet121 features both dominating top rankings.

The smartphone-based implementation might eliminate barriers constraining adoption of digital PB monitoring. Unmanned aerial vehicles require substantial financial investment, specialised piloting expertise, and regulatory compliance [72,74], whilst satellite methods demand advanced image processing and face temporal constraints from cloud cover [37,75]. By achieving accuracy approaching satellite systems using ubiquitous devices, this work positioned smartphone imaging as potentially transformative technology for precision grazing management. However, severe LOFO-CV degradation indicated widespread adoption would require either expanded multi-farm training datasets capturing greater environmental diversity or efficient transfer learning protocols enabling rapid farm-specific adaptation.

4.6. Limitations and Future Research Directions

This study faces several methodological constraints affecting interpretation and generalisation. Single images captured per paddock cannot fully represent spatial heterogeneity, as PB exhibits substantial variation from soil heterogeneity, topographic moisture gradients, differential grazing pressure, and proximity to water points. The absence of spatial coordinates for individual RPM measurement points prevented exploitation of within-paddock spatial variability through point-level prediction approaches that could leverage geotagged smartphone imagery for fine-scale biomass mapping. Multiple sampling locations per paddock could improve spatial representation but introduce labour constraints limiting scalability.

Illumination variability and pre-processing trade-offs. The decision to preserve natural illumination characteristics without explicit normalisation represents a trade-off between spectral fidelity and robustness. Whilst this approach allowed vegetation indices and CNN features to learn from naturally varying lighting conditions across seasons and times of day, it introduces systematic errors under extreme illumination scenarios (e.g., harsh shadows, overcast skies). Future implementations could explore adaptive illumination correction methods such as Retinex algorithms [76] or grey-world assumptions [51] to normalise colour consistency whilst preserving biologically relevant spectral information. Similarly, the decision to avoid training data augmentation prioritised preservation of authentic pasture spectral signatures over synthetic dataset expansion. However, moderate augmentation strategies (For example, brightness adjustment ±10%, horizontal flipping) informed by empirical studies [77,78] could improve model robustness without substantially compromising spectral fidelity, particularly for improving generalisation to unseen farm environments as revealed by LOFO-CV results.

Furthermore, RPM ground truth measurements contain inherent uncertainty, with calibration relationships typically explaining 70-80% of variance [45], establishing a fundamental ceiling on achievable prediction accuracy. Training dataset size of 3291 images, whilst sufficient for TL approaches, proved inadequate for training DL architectures from scratch, suggesting that expanding to tens of thousands of images may enable end-to-end fine-tuning whilst improving spatial generalisation. Temporal representation remains limited when relying on single-date imagery that cannot capture growth trajectory dynamics.

Future research directions should prioritise several complementary advances. Recurrent neural network architectures incorporating LSTM layers could exploit temporal image sequences to predict biomass trajectories, potentially improving forecast accuracy for grazing management planning. Attention mechanisms embedded within CNN architectures may automatically identify biomass-relevant image regions, improving interpretability and prediction robustness. Uncertainty quantification through Bayesian deep learning or ensemble variance estimation would provide prediction confidence intervals supporting risk-based decision-making. Three-dimensional imaging approaches using stereo cameras or structure-from-motion photogrammetry may improve estimation for prostrate species like kikuyu by capturing canopy height information unavailable from RGB imagery. Automated RPM systems and sensor-assisted calibration pipelines represent an important frontier, enabling continuous or semi-autonomous measurement streams that enrich training datasets and strengthen model accuracy across seasons. These advances require continued expansion of large, diverse and representative training datasets, as model performance ultimately depends on exposure to sufficient variability in pasture conditions, environmental drivers and management practices. Most critically, on-farm validation trials remain essential for validating practical utility, assessing predictive accuracy, evaluating user acceptance and workflow integration, and quantifying economic returns.

5. Conclusion

This study demonstrates accurate automated pasture biomass estimation from smartphone imagery through systematic integration of transfer learning, ensemble methods, and weather data. Evaluating 21 modelling configurations across 3291 images from 15 dairy farms, stacking ensemble methods combining pre-trained convolutional neural network features with hand-crafted vegetation indices and weather variables achieved $R^2 = 0.561$, MAE = 351 kg DM ha^{-1}, and RMSE = 437 kg DM ha^{-1}. Transfer learning using frozen ImageNet features dramatically outperformed deep learning models trained from scratch, confirming that pre-trained visual representations provide robust feature extraction for pasture with moderate training data. Multi-architecture feature integration captured complementary information, with DenseNet121 features dominating mutual information rankings alongside hand-crafted vegetation indices. Stacking ensemble with Ridge regression meta-learner achieved optimal integration through learned combination weights, whilst weather data integration provided 5.9% performance improvement by incorporating meteorological variables beyond image-derived features. Seasonal analysis revealed optimal spring performance coinciding with peak ryegrass growth, whilst winter exhibited reduced accuracy during species transition periods. However, leave-one-farm-out cross-validation exposed severe spatial generalisation constraints, with mean R^2 approaching zero despite strong within-dataset performance, indicating farm-specific dependencies limit zero-shot transfer to new environments. Future research priorities should consider recurrent neural network architectures for temporal sequence modelling, attention mechanisms for automated region identification, uncertainty quantification for decision support, expanded multi-farm training datasets capturing greater environmental diversity, and on-farm trials assessing real-world feasibility in

commercial dairy farm management systems.

CRediT authorship contribution statement

Blessing Nnenna Azubuike: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Chlingaryan:** Writing – review & editing, Supervision, Conceptualization. **Martin Correa-Luna:** Writing – review & editing, Supervision, Conceptualization. **Cameron E.F. Clark:** Writing – review & editing, Supervision, Conceptualization. **Sergio C. Garcia:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition, Conceptualization.

Funding statement

This research was funded by the Dairy UP program, a collaborative RD&E program for New South Wales, Australia (www.dairyup.com.au) through the academic scholarship awarded to Blessing N. Azubuike.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the research scholarship granted by the DairyUP project at the University of Sydney.

Appendix

Table A1
Summary of 21 experimental configurations evaluated for pasture biomass prediction

Experiment ID	Phase	Model(s)	Features	Key Parameters	Purpose
Baseline Models					
1	Baseline	Random Forest	Hand-crafted (67)	200 trees, max depth 20	Baseline without DL
2	Baseline	XGBoost	Hand-crafted (67)	200 trees, max depth 7, LR 0.1	Baseline without DL
3	Baseline	Gradient Boosting	Hand-crafted (67)	200 trees, max depth 5, LR 0.1	Baseline without DL
Deep Learning from Scratch					
4	DL Scratch	ResNet50 (fine-tuned)	Raw images	50 epochs, BS 16, LR 0.0001	End-to-end learning
5	DL Scratch	Custom Attention CNN	Raw images	Spatial + channel attention	Custom architecture
6	DL Scratch	Density Estimation Network	Raw images	Fully convolutional, 56 × 56	Pixel-level prediction
7	DL Scratch	Vision Transformer	Raw images	8 encoder layers, 196 patches	Transformer approach
Transfer Learning (Individual Architectures)					
8	Transfer Learning	RF on ResNet50	ResNet50 frozen (2,048)	200 trees, max depth 20	Single CNN evaluation
9	Transfer Learning	RF on VGG16	VGG16 frozen (512)	200 trees, max depth 20	Single CNN evaluation
10	Transfer Learning	RF on MobileNetV2	MobileNetV2 frozen (1,280)	200 trees, max depth 20	Single CNN evaluation
11	Transfer Learning	RF on DenseNet121	DenseNet121 frozen (1,024)	200 trees, max depth 20	Single CNN evaluation
Multi-CNN Ensembles					
12	Multi-CNN	RF, XGBoost, GBM averaged	Multi-CNN (4,864) + Hand-crafted (67) + Metadata (20) → 120 selected	Initial parameters, equal weights	Multi-architecture integration
13	Multi-CNN	RF, XGBoost, GBM averaged	Exp. 12 + Weather (8) → 120 selected	Initial parameters, equal weights	Weather data contribution
14	Hyperparameter Optimisation	RF, XGBoost, GBM averaged	Same as Exp. 13	RandomizedSearchCV, 50 iterations	Optimised parameters
Advanced Ensemble Methods					
15	Ensemble	Weighted ensemble	Same as Exp. 14	SLSQP optimised weights	Optimised combination
16	Ensemble	Stacking (no weather)	Multi-CNN (4,864) + Hand-crafted (67) + Metadata (20) → 120 selected	Ridge meta-learner, α = 1.0	Stacking without weather
17	Ensemble	Stacking (with weather)	Exp. 16 + Weather (8) → 120 selected	Ridge meta-learner, α = 1.0	Best model
Data Augmentation					
18	Augmentation	RF on ResNet50	ResNet50 frozen, augmented training (× 3)	Aug: rotation ±20°, flip, shift ±10%, brightness 0.8-1.2	Synthetic data expansion
Grass-Specific Models					
19	Grass-specific	Stacking (ryegrass only)	Multi-CNN + Hand-crafted + Metadata + Weather → 80 selected	May-Nov data (n = 2058), Ridge meta-learner	Species-specific model
20	Grass-specific	Stacking (kikuyu only)	Multi-CNN + Hand-crafted + Metadata + Weather → 80 selected	Dec-Apr data (n = 1231), Ridge meta-learner	Species-specific model
21	Grass-specific	Combined grass ensemble	Route to Exp. 19 or 20 by month	Same as Exp. 17	Species-adaptive routing

LR = learning rate; BS = batch size; Aug = augmentation; DL = deep learning; CNN = convolutional neural network; RF = Random Forest; GBM = Gradient Boosting Machine. Multi-CNN features comprise ResNet50 (2048), VGG16 (512), MobileNetV2 (1280), and DenseNet121 (1024). Feature selection employed mutual information regression implemented through SelectKBest.

Table A2
Comparative performance of 18 modelling approaches ranked by R^2

Rank	Experiment	Model	Category	R^2	RMSE (kg DM ha^{-1})	MAE (kg DM ha^{-1})	Notes
1	17	Stacking + Weather	Advanced Ensemble Methods	0.561	437	351	Best model
2	15	Weighted Ensemble	Advanced Ensemble Methods	0.549	442	355	Grid search weights
3	14	Hyperparameter Optimised Ensemble	Hyperparameter Optimisation	0.538	448	361	RandomizedSearchCV
4	16	Stacking No Weather	Advanced Ensemble Methods	0.530	451	364	Ridge meta-learner
5	13	Multi-CNN + Weather	Multi-CNN Ensembles	0.521	456	368	Simple averaging
6	12	Multi-CNN Ensemble	Multi-CNN Ensembles	0.495	468	378	No weather
7	11	DenseNet121 Transfer Learning	Transfer Learning	0.450	489	390	Single CNN
8	2	XGBoost Baseline	Baseline Models	0.380	519	410	Hand-crafted features only
9	3	Gradient Boosting Baseline	Baseline Models	0.369	523	412	Hand-crafted features only
10	1	Random Forest Baseline	Baseline Models	0.357	528	421	Hand-crafted features only
11	9	VGG16 Transfer Learning	Transfer Learning	0.347	532	426	Single CNN
12	10	MobileNetV2 Transfer Learning	Transfer Learning	0.331	539	431	Single CNN
13	6	Density Estimation Network	Deep Learning from Scratch	0.223	581	468	No spatial labels
14	8	ResNet50 Transfer Learning	Transfer Learning	0.152	606	495	Single CNN
15	18	Training Augmentation	Data Augmentation	0.140	611	498	Feature mismatch
16	4	End-to-End ResNet50	Deep Learning from Scratch	0.010	655	538	Prediction collapse
17	7	Vision Transformer	Deep Learning from Scratch	0.007	656	526	Requires large dataset
18	5	Custom Attention CNN	Deep Learning from Scratch	−0.781	879	735	Insufficient data

All experiments used identical train-test splits (2632 training samples, 659 test samples, random state 42). Performance metrics computed on held-out test set. Grass-specific models (Experiments 19-21) excluded from table as they employed separate seasonal train-test partitions. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error.

Data availability

Data will be made available on request.

References

[1] O. Dubovik, G.L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, Z. Li, Grand Challenges in Satellite Remote Sensing [Specialty Grand Challenge], Front. Rem. Sens. 2 (2021), https://doi.org/10.3389/frsen.2021.619818, 2021.

[2] S. Gard, M. Neal, E. Minnee, Pasture performance tools: current and future state, J. New Zealand Grasslands 86 (2024) 273–279, https://doi.org/10.33584/jnzg.2024.86.3688.

[3] S. Ghajar, B. Tracy, Proximal sensing in grasslands and pastures, Agriculture 11 (8) (2021) 740, https://doi.org/10.3390/agriculture11080740.

[4] D.J. Murphy, M.D. Murphy, B. O'Brien, M. O'Donovan, A review of precision technologies for optimising pasture measurement on Irish grassland, Agriculture 11 (7) (2021) 600, https://doi.org/10.3390/agriculture11070600.

[5] R. Kallenbach, S. Hamilton, T.R. Lock, 345 digital technologies ease the burden of pasture intake measurements: a new hope, J. Anim. Sci. 98 (Supplement_4) (2020), https://doi.org/10.1093/jas/skaa278.142, 78-78.

[6] V. Antognoli, L. Presutti, M. Bovo, D. Torreggiani, P. Tassinari, Computer vision in dairy farm management: a literature review of current applications and future perspectives, Animals 15 (17) (2025) 2508, https://doi.org/10.3390/ani15172508.

[7] R.E.P. Ferreira, J.R.R. Dórea, em>International symposium on ruminant physiology: leveraging computer vision, large language models, and multimodal machine learning for optimal decision making in dairy farming, J. Dairy Sci. 108 (7) (2025) 7493–7510, https://doi.org/10.3168/jds.2024-25650.

[8] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: a survey, Comput. Electron. Agric. 147 (2018) 70–90, https://doi.org/10.1016/j.compag.2018.02.016.

[9] M. El Sakka, M. Ivanovici, L. Chaari, J. Mothe, A review of CNN applications in smart agriculture using multimodal data, Sensors 25 (2) (2025) 472, https://doi.org/10.3390/s25020472.

[10] M. El Sakka, J. Mothe, M. Ivanovici, Images and CNN applications in smart agriculture, European J. Rem. Sens. 57 (1) (2024) 2352386, https://doi.org/10.1080/22797254.2024.2352386.

[11] E. Elbasi, C. Zaki, A.E. Topcu, W. Abdelbaki, A.I. Zreikat, E. Cina, A. Shdefat, L. Saker, Crop prediction model using machine learning algorithms, Appl. Sci. 13 (16) (2023) 9288, https://doi.org/10.3390/app13169288.

[12] Z. Al Sahili, M. Awad, The power of transfer learning in agricultural applications: AgriNet [Original Research], Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.992700, 2022.

[13] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A.S. Albahri, B.S.N. Al-dabbagh, M.A. Fadhel, M. Manoufali, J. Zhang, A.H. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, Y. Gu, A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications, J. Big Data 10 (1) (2023) 46, https://doi.org/10.1186/s40537-023-00727-2.

[14] M. Hossen, M. Awrangjeb, S. Pan, A. Mamun, Transfer learning in agriculture: a review, Artif. Intell. Rev. 58 (2025), https://doi.org/10.1007/s10462-024-11081-x.

[15] I. Adhikari, B. Koirala, M.R. Panda, Decoding plant diseases: unleashing the power of ResNet-50 for superior disease detection, in: 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), 2024, 27-29 Jan. 2024.

[16] U. Arora, U. Mishra, S. Singh, V. Singh, Comparative analysis of VGG16, inception V4, AlexNet, and ResNet 50 for plant disease identification, in: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024, 24-28 June 2024.

[17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 27-30 June 2016.

[18] G. Pei, X. Qian, B. Zhou, Z. Liu, W. Wu, Research on agricultural disease recognition methods based on very large kernel convolutional network-RepLKNet, Sci. Rep. 15 (1) (2025) 16843, https://doi.org/10.1038/s41598-025-01553-7.

[19] S.R. Shah, S. Qadri, H. Bibi, S.M.W. Shah, M.I. Sharif, F. Marinello, Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease, Agronomy 13 (6) (2023) 1633, https://doi.org/10.3390/agronomy13061633.

[20] N. Singh, Plant diseases detection using deep learning and machine vision, in: 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2024, 14-16 March 2024.

[21] C. Zheng, A. Abd-Elrahman, V.M. Whitaker, C. Dalid, Deep learning for strawberry canopy delineation and biomass prediction from high-resolution images, Plant Phenom. 2022 (2022) 9850486, https://doi.org/10.34133/2022/9850486.

[22] B. Cândido, U. Mindala, H. Ebrahimy, Z. Zhang, R. Kallenbach, Integrating proximal and remote sensing with machine learning for pasture biomass estimation, Sensors 25 (7) (2025) 1987. https://www.mdpi.com/1424-8220/25/7/1987.

[23] M.G. Ogungbuyi, J. Guerschman, A.M. Fischer, R.A. Crabbe, I. Ara, C. Mohammed, P. Scarth, P. Tickle, J. Whitehead, M.T. Harrison, Improvement of pasture biomass modelling using high-resolution satelli te imagery and machine learning,

J. Environ. Manag. 356 (2024) 120564, https://doi.org/10.1016/j.jenvman.2024.120564.
[24] M. Correa-Luna, J. Gargiulo, P. Beale, D. Deane, J. Leonard, J. Hack, Z. Geldof, C. Wilson, S. Garcia, Accounting for minimum data required to train a machine learning model to accurately monitor Australian dairy pastures using remote sensing, Sci. Rep. 14 (1) (2024), https://doi.org/10.1038/s41598-024-68094-3.
[25] J. Neal, W. Fulkerson, B. Sutton, Differences in water-use efficiency among perennial forages used by the dairy industry under optimum and deficit irrigation, Irrig. Sci. 29 (2011) 213–232, https://doi.org/10.1007/s00271-010-0229-1.
[26] C.O.G. Bazzo, B. Kamali, C. Hütt, G. Bareth, T. Gaiser, A review of estimation methods for aboveground biomass in grasslands using UAV, Remote Sens. 15 (3) (2023) 639, https://doi.org/10.3390/rs15030639.
[27] F. Dhawi, A. Ghafoor, N. Almousa, S. Ali, S. Alqanbar, Predictive modelling employing machine learning, convolutional neural networks (CNNs), and smartphone RGB images for non-destructive biomass estimation of pearl millet (Pennisetum glaucum) [Original Research], Front. Plant Sci. 16 (2025), https://doi.org/10.3389/fpls.2025.1594728, 2025.
[28] V. Franco, M. Hott, R. Andrade, L. Goliatt, Hybrid machine learning methods combined with computer vision approaches to estimate biophysical parameters of pastures, Evolution. Intell. 16 (2022), https://doi.org/10.1007/s12065-022-00736-9.
[29] A. Michez, P. Lejeune, S. Bauwens, A.A.L. Herinaina, Y. Blaise, E. Castro Muñoz, F. Lebeau, J. Bindelle, Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system, Remote Sens. 11 (5) (2019) 473. https://www.mdpi.com/2072-4292/11/5/473.
[30] R. Näsi, N. Viljanen, J. Kaivosoja, K. Alhonoja, T. Hakala, L. Markelin, E. Honkavaara, Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features, Remote Sens. 10 (7) (2018) 1082, https://doi.org/10.3390/rs10071082.
[31] M.T. Schaefer, D.W. Lamb, A combination of plant NDVI and LiDAR measurements improve the estimation of pasture biomass in tall fescue (Festuca arundinacea var. fletcher), Remote Sens. 8 (2) (2016) 109, https://doi.org/10.3390/rs8020109.
[32] L. Woodrow, J. Carter, G. Fraser, J. Barnetson, Using continuous output neural nets to estimate pasture biomass from digital photographs in grazing lands, AgriEngineering 5 (2) (2023) 1051–1067, https://doi.org/10.3390/agriengineering5020066.
[33] A.A. Dos Reis, J.P.S. Werner, B.C. Silva, G.K.D.A. Figueiredo, J.F.G. Antunes, J.C.D. M. Esquerdo, A.C. Coutinho, R.A.C. Lamparelli, J.V. Rocha, P.S.G. Magalhães, Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery, Remote Sens. 12 (16) (2020) 2534. https://www.mdpi.com/2072-4292/12/16/2534.
[34] R.G. Freitas, F.R.S. Pereira, A.A. Dos Reis, P.S.G. Magalhães, G.K.D.A. Figueiredo, L.R. do Amaral, Estimating pasture aboveground biomass under an integrated crop-livestock system based on spectral and texture measures derived from UAV images, Comput. Electron. Agric. 198 (2022) 107122, https://doi.org/10.1016/j.compag.2022.107122.
[35] L. Zhang, B. Verma, D. Stockwell, S. Chowdhury, Density weighted connectivity of grass pixels in image frames for biomass estimation, Expert Syst. Appl. 101 (2018) 213–227, https://doi.org/10.1016/j.eswa.2018.01.055.
[36] J. Ali, W. Haoran, K. Mehmood, W. Hussain, F. Iftikhar, F. Shahzad, K. Hussain, Y. Qun, J. Zhongkui, Remote sensing and integration of machine learning algorithms for above-ground biomass estimation in Larix principis-rupprechtii Mayr plantations: a case study using Sentinel-2 and Landsat-9 data in northern China [Original Research], Front. Environ. Sci. 13 (2025), https://doi.org/10.3389/fenvs.2025.1577298.
[37] Y. Chen, J. Guerschman, Y. Shendryk, D. Henry, M.T. Harrison, Estimating pasture biomass using Sentinel-2 imagery and machine learni ng, Remote Sens. 13 (4) (2021) 603, https://doi.org/10.3390/rs13040603.
[38] H. Jin, Y. Zhao, U. Pak, Z. Zhen, K. So, Assessing the effect of ensemble learning algorithms and validation approach on estimating forest aboveground biomass: a case study of natural secondary forest in Northeast China, Geo-Spatial Inf. Sci. 28 (2) (2025) 609–628, https://doi.org/10.1080/10095020.2024.2311261.
[39] M. Luo, S.A. Anees, Q. Huang, X. Qin, Z. Qin, J. Fan, G. Han, L. Zhang, H.Z. M. Shafri, Improving forest above-ground biomass estimation by integrating individual machine learning models, Forests 15 (6) (2024) 975, https://doi.org/10.3390/f15060975.
[40] Y. Zhang, J. Ma, S. Liang, X. Li, J. Liu, A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets, GIScience Remote Sens. 59 (1) (2022) 234–249, https://doi.org/10.1080/15481603.2021.2023842.
[41] F. Shahzad, K. Mehmood, S. Anees, M. Adnan, S. Muhammad, I. Haidar, J. Ali, K. Hussain, Z. Feng, W. Khan, Advancing forest fire prediction: a multi-layer stacking ensemble model approach, Earth Sci. Inform. 18 (2025) 270, https://doi.org/10.1007/s12145-025-01782-4.
[42] J. Bendig, K. Yu, H. Aasen, A. Bolten, S. Bennertz, J. Broscheit, M.L. Gnyp, G. Bareth, Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley, Int. J. Appl. Earth Obs. Geoinf. 39 (2015) 79–87, https://doi.org/10.1016/j.jag.2015.02.012.
[43] M. Ashfaq, I. Khan, R. Afzal, D. Shah, E.D.S. Ali Aurakzai, M. Tahir, Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques, Sci. Rep. 15 (2025) 25, https://doi.org/10.1038/s41598-025-02700-w.
[44] T. Morais, R. Teixeira, T. Domingos, A spatially explicit life cycle assessment midpoint indicator for soil quality in the European Union using soil organic carbon, Int. J. Life Cycle Assess. 21 (2016), https://doi.org/10.1007/s11367-016-1077-x.
[45] D.F. Earle, A.A. McGowan, Evaluation and calibration of an automated rising plate meter for estimating dry matter yield of pasture, Aust. J. Exp. Agric. 19 (1979) 337–343, https://doi.org/10.1071/EA9790337.
[46] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, Segformer: Simple and Efficient Design for Semantic Segmentation with Transformers, 2021.
[47] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Álvarez, P. Luo, Segformer: Simple and Efficient Design for Semantic Segmentation with Transformers, Neural Information Processing Systems, 2021.
[48] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[49] I. Yeuseyenka, I. Melnikau, I. Yemelyanov, Detection and selection of moving objects in video images based on impulse and recurrent neural networks, J. Data Anal. Inf. Process. 10 (2022) 127–141, https://doi.org/10.4236/jdaip.2022.102008.
[50] J. Xue, B. Su, Significant remote sensing vegetation indices: a review of developments and applications, J. Sens. 2017 (1) (2017) 1353691, https://doi.org/10.1155/2017/1353691.
[51] J.v. d. Weijer, T. Gevers, A. Gijsenij, Edge-based color constancy, IEEE Trans. Image Process. 16 (9) (2007) 2207–2214, https://doi.org/10.1109/TIP.2007.901808.
[52] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621, https://doi.org/10.1109/TSMC.1973.4309314.
[53] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987, https://doi.org/10.1109/TPAMI.2002.1017623.
[54] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations (ICLR 2015), 2015. San Diego, 7-9 May 2015. arXiv:1409.1556.
[55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, 18-23 June 2018.
[56] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 21-26 July 2017.
[57] S. Edwards, Elements of Information Theory, second ed., 2008.
[58] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[59] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: KDD '16: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.
[60] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[61] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, Neural Inform. Process. Sys. (2017), https://doi.org/10.48550/arXiv.1706.03762.
[62] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module, Computer Vision – ECCV 2018, Cham, 2018.
[63] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2020. *ArXiv, abs/2010.11929*.
[64] J. Heaton, in: Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep Learning, The MIT Press, 2017, p. 800, https://doi.org/10.1007/s10710-017-9314-z, 2016, ISBN: 0262035618. Genetic Programming and Evolvable Machines, 19.
[65] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? ArXiv, abs/1411.1792 (2014).
[66] J.I. Gargiulo, N.A. Lyons, F. Masia, P. Beale, J.R. Insua, M. Correa-Luna, S. C. Garcia, Comparison of ground-based, unmanned aerial vehicles and satellite rem ote sensing technologies for monitoring pasture biomass on dairy farms, Remote Sens. 15 (11) (2023) 2752, https://doi.org/10.3390/rs15112752.
[67] A.R. Lawson, K. Giri, A.L. Thomson, S.B. Karunaratne, K.F. Smith, J.L. Jacobs, E. M. Morse-McNabb, Multi-site calibration and validation of a wide-angle ultrasonic senso r and precise GPS to estimate pasture mass at the paddock scale, Comput. Electron. Agric. 195 (2022) 106786, https://doi.org/10.1016/j.compag.2022.106786.
[68] L.A. Villalobos-Villalobos, R. WingChing-Jones, Forage biomass estimated with a pre-calibrated equation of a rising platemeter in pastures grown in tropical conditions, Grasses 2 (2) (2023) 127–141, https://doi.org/10.3390/grasses2020011.
[69] S. García, M. Islam, C. Clark, P. Martin, Kikuyu-based pasture for dairy production: a review, Crop Pasture Sci. 65 (8) (2014) 787–797, https://doi.org/10.1071/CP13414.
[70] L.K. Vanamburg, M.J. Trlica, R.M. Hoffer, M.A. Weltz, Ground based digital imagery for grassland biomass estimation, Int. J. Rem. Sens. 27 (5) (2006) 939–950, https://doi.org/10.1080/01431160500114789.
[71] S. Taugourdeau, A. Diedhiou, C. Fassinou, M. Bossoukpe, O. Diatta, A. N'Goran, A. Auderbert, O. Ndiaye, A.A. Diouf, T. Tagesson, R. Fensholt, E. Faye, Estimating herbaceous aboveground biomass in Sahelian rangelands using structure from motion data collected on the ground and by UAV, Ecol. Evol. 12 (5) (2022) e8867, https://doi.org/10.1002/ece3.8867.

[72] M. Vahidi, S. Shafian, S. Thomas, R. Maguire, Pasture biomass estimation using ultra-high-resolution RGB UAVs images and deep learning, Remote Sens. 15 (24) (2023) 5714, https://doi.org/10.3390/rs15245714.
[73] M. Colovic, A.M. Stellacci, N. Mzid, M. Di Venosa, M. Todorovic, V. Cantore, R. Albrizio, Comparative performance of aerial RGB vs. ground hyperspectral indices for evaluating water and nitrogen status in sweet maize, Agronomy 14 (3) (2024) 562, https://doi.org/10.3390/agronomy14030562.
[74] U. Lussem, A. Bolten, J. Menne, M. Gnyp, J. Schellberg, G. Bareth, Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices, J. Appl. Remote Sens. 13 (2019) 1, https://doi.org/10.1117/1.JRS.13.034525.
[75] A. Edirisinghe, D. Clark, D. Waugh, Spatio-temporal modelling of biomass of intensively grazed perennial d airy pastures using multispectral remote sensing, Int. J. Appl. Earth Obs. Geoinf. 16 (2012) 5–16, https://doi.org/10.1016/j.jag.2011.11.006.
[76] D.J. Jobson, Z. Rahman, G.A. Woodell, Properties and performance of a center/surround retinex, IEEE Trans. Image Process. 6 (3) (1997) 451–462, https://doi.org/10.1109/83.557356.
[77] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning. https://doi.org/10.48550/arXiv.1712.04621, 2017.
[78] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 60, https://doi.org/10.1186/s40537-019-0197-0.

CHAPTER 8

General Discussion and Future Directions

Artificial intelligence (AI) applications in pasture-based dairy systems have demonstrated capability in controlled research settings, but practical implementation across diverse commercial farm environments with quantifiable production outcomes remained limited. This thesis developed and validated AI methods across precision feeding optimisation, pasture biomass (PB) estimation, and automated grazing detection. The preceding chapters demonstrated that these components function independently with measurable performance, and establishing technical feasibility for commercial application. Building on these foundations, this chapter consolidates and integrates the major findings, discusses implementation pathways, acknowledges limitations, and proposes future research directions for enhancing production efficiency in pasture-based dairy systems (P-BDS).

KEY RESEARCH OUTCOMES

In Chapter 3, we examined whether ML combined with evolutionary algorithms could leverage individual cow variability to increase milk production without additional feed costs. We integrated Random Forest predictions with differential evolution algorithms across 81 cows over 91 days, observing an 8% theoretical increase in herd-level daily milk yield without increasing concentrate allocation. Building on these promising results, Chapter 4 implemented the approach on a commercial farm with 165 cows over 30 days, evaluating four multi-objective evolutionary algorithms (NSGA-II, SPEA-II, SMS-EMOA, RVEA) to compare their computational efficiency and production outcomes. Statistical validation across ten independent runs demonstrated MY increases ranging from 6.63% to 8.64% depending on algorithm selection. NSGA-II achieved the highest performance (8.64%) with fastest computation time, whilst SPEA-II delivered competitive performance (7.94%). These findings address the critical research gap identified in the literature review (Cockburn, 2020; Dela Rue & Eastwood, 2017) and provide first data-driven predictive evidence that automated systems integrating ML with optimisation algorithms can estimate potential production gains under commercial conditions by exploiting cow-to-cow variability for individualised supplementation decisions. Whilst mechanistic equations such as those detailed in the NRC (2001) can theoretically be applied at the individual cow level, they are calibrated on population averages and do not adapt to the specific conditions, pasture base, or biological variability of individual farm systems. The data-driven approach developed in Chapters 3 and 4 addresses this limitation directly by learning directly from farm-specific production records, capturing system-specific relationships between cow characteristics, concentrate allocation, and milk yield that population-derived equations cannot adequately represent.

Pasture biomass estimation was targeted for modelling in Chapter 5 and it required a different strategy due to different reasons. We developed two complementary approaches to address the critical limitations of remote sensing satellites identified in the literature review including cloud interference, vegetation index saturation, and spatial transferability across diverse farm environments. In Chapter 5, satellite-based methods using Sentinel-2 imagery mitigated temporal data gaps caused by cloud interference through multiquadric radial basis interpolation of rising plate meter measurements, whilst addressing vegetation index saturation by using full spectral bands rather than vegetation indices across 16 farms. We observed that augmenting training datasets with approximately 30% interpolated observations improved performance from baseline R^2 of 0.63 (MAE = 243 kg DM/ha, RMSE = 313 kg DM/ha) to R^2 of 0.70 (MAE = 216 kg DM/ha, RMSE = 293 kg DM/ha), achieving comparable accuracy to commercial platforms whilst using freely available satellite data.

For Chapter 6, we developed ML frameworks for automated grazing event detection from satellite data across 12 commercial dairy farms using the interpolated data developed from Chapter 5. We observed

that Random Forest achieved within-year detection performance of F1 = 0.878 (precision = 0.938, recall = 0.825), demonstrating that temporal PB patterns from interpolated satellite observations provide robust signals for identifying grazing events. However, when we tested these models across production years, One-Class Support Vector Machine demonstrated superior temporal transferability with F1 = 0.692 (precision = 0.610, recall = 0.799), outperforming supervised models by 7.6% on independent Year 2 data where supervised approaches experienced 24.2% average performance degradation. Independent validation on 218 GPS-confirmed grazing events demonstrated strong pasture utilisation quantification accuracy (pre-grazing R^2 = 0.966, post-grazing R^2 = 0.998, removal R^2 = 0.922), establishing that satellite systems accurately quantify PB availability, residuals, and removal when adequate temporal alignment exists between measurements and grazing events. Complementing these satellite-based approaches, Chapter 7 evaluated image-based PB estimation from smartphone imagery paired with rising plate meter ground truth data, using transfer learning with frozen convolutional neural network features and stacking ensemble methods. We achieved R^2 of 0.561 (MAE = 351 kg DM/ha, RMSE = 437 kg DM/ha), with multi-architecture integration combining DenseNet121, ResNet50, VGG16, and MobileNetV2 features and capturing complementary visual patterns. Critically, smartphone methods maintained sensitivity across broader PB ranges where satellite-derived vegetation indices saturate, positioning these approaches as complementary rather than competing technologies, though substantial LOFO-CV performance degradation indicates that generalisation to unseen farm environments remains a critical limitation requiring farm-specific calibration data to resolve.

Collectively, these findings address the aim of this thesis of developing and validating AI methods for precision feeding optimisation, PB estimation, and automated grazing detection in P-BDS. This thesis makes three distinct contributions extending scientific understanding beyond what was previously known. First, the precision feeding work provides first empirical evidence that ML-based algorithms integrated with evolutionary algorithms for individualised feeding quantifies production gains under commercial conditions, moving beyond theoretical correlational models (Objective 1). The key innovation lies in embedding predictive models directly within optimisation frameworks to exploit individual cow variability, enabling exploration of supplementation strategies that conventional rule-based approaches cannot evaluate whilst maintaining computational tractability for commercial deployment. Second, it develops novel methodological approaches for overcoming persistent limitations in PB estimation. The PB estimation work demonstrates that multiquadric interpolation overcomes temporal data gaps from cloud interference when using satellites to monitor pastures, with progressive training maintaining model accuracy across seasons by continuously updating parameters with new observations rather than relying solely on historical data. Transfer learning from smartphone imagery provides an accessible (albeit not achieving enough accuracy yet) alternative for ground-level PB estimation, demonstrating that pretrained features from general computer vision combined with ensemble methods can compensate for limited domain-specific training data (Objective 2). For satellites

facing cloud interference and limited revisit frequency, the innovation combines data augmentation with adaptive learning to maintain accuracy despite incomplete temporal coverage. For ground-based systems with limited training data, the innovation demonstrates that knowledge transfer from general computer vision can compensate for scarcity of domain-specific data. This establishes that platform selection and ML methodology must be matched to address specific constraints, with each platform offering complementary strengths for comprehensive farm monitoring.

Third, it establishes proof-of-concept for automated grazing event detection using freely available satellite data, demonstrating that temporal PB patterns contain sufficient signal for reliable detection whilst revealing fundamental trade-offs between absolute accuracy and temporal transferability. The grazing event detection work establishes that semi-supervised anomaly detection outperforms fully supervised approaches for cross-year transferability where interannual variability constrains model performance (Objective 3). The superior cross-year performance of semi-supervised anomaly detection compared to fully supervised classification approaches provides practical guidance for researchers and practitioners developing automated monitoring systems where training data availability constrains model development. This finding challenges common assumptions that supervised learning always outperforms alternative paradigms when labelled training data exists, revealing that semi-supervised approaches learning normal pasture growth patterns can be robust to interannual variation in environmental conditions. This provides practical guidance for developing automated monitoring systems where collecting extensive labelled data across multiple years proves prohibitively expensive, suggesting that anomaly detection frameworks may achieve better long-term reliability than classification approaches for agricultural applications experiencing high environmental variability between seasons. However, whilst individual components demonstrate technical feasibility and address critical research gaps, we identified some limitations across the three domains that warrant careful consideration for future research.

LIMITATIONS AND RESEARCH GAPS

First, the precision feeding optimisation studies (Chapters 3 and 4) achieved demonstrable performance improvements, but the results represent theoretical potential derived from correlational data within existing management practices rather than causal evidence from experimental manipulation of concentrate allocation. The dataset reflects management decisions based on existing production levels, describing associations under standard practice rather than direct cause-and-effect responses to systematic feed rate changes. This fundamental constraint prevents definitive claims about whether the observed 8% to 8.5% yield improvements would materialise under controlled experimental conditions where concentrate allocation is actively manipulated according to model recommendations. Insufficient training data per individual cow further limited the ability of models to learn effectively from historical

patterns, whilst fixed feeding rates restricted exploration of diverse supplementation strategies. Validation through controlled field trials with systematic concentrate manipulation remains essential to confirm practical efficacy and quantify actual economic returns before commercial deployment can be recommended with confidence.

Second, satellite-based PB estimation in Chapter 5 encountered temporal and spatial transferability constraints that affect practical deployment (Meyer & Pebesma, 2021; Ogungbuyi et al., 2024). Temporal generalisation degraded substantially when models trained on Year 1 data were tested on Year 2, with performance declining to R^2 of 0.24 compared to within-year accuracy of $R^2 = 0.70$. Independent validation sets achieved moderate performance ($R^2 = 0.44$ and $R^2 = 0.48$), demonstrating proof-of-concept but indicating that substantially more diverse training data spanning multiple years and environmental conditions would be required to achieve the robustness necessary for reliable commercial application. Multiquadric interpolation, whilst effective at filling cloud-induced data gaps, may smooth abrupt PB declines following intensive grazing events, potentially underestimating utilisation rates during rapid grazing periods. Additional, daily weather aggregates used in the models may overlook short-lived environmental events that influence pasture growth dynamics, introducing temporal resolution mismatches between predictor variables and actual biological processes. Furthermore, all data-driven approaches remain fundamentally reliant on obtaining accurate ground truth measurements, which are expensive, time-consuming to collect, and require continuous quality verification to ensure AI systems produce acceptable outcomes. This dependency on labour-intensive field measurements creates a persistent bottleneck constraining the scalability and reliability of machine learning-based monitoring systems, emphasising the critical need for automated ground truth collection infrastructure discussed in future research directions.

Third, grazing event detection in Chapter 6 faced the most severe temporal resolution limitation, with 80.6% of GPS-confirmed events excluded from validation due to insufficient temporal alignment between satellite observations and grazing periods. Of 1,492 documented grazing events, only 218 had adequate temporal bracketing with satellite measurements, fundamentally constraining our ability to validate the detection framework across the full spectrum of management scenarios. This severe exclusion rate prevented comprehensive reporting of yearly pasture utilisation at both farm and paddock levels, limiting our capacity to quantify cumulative grazing patterns and establish baseline utilisation metrics for commercial dairy farms. The 5-day nominal revisit frequency of Sentinel-2 combined with cloud interference creates temporal gaps that miss the 24 to 48 hour grazing periods characteristic of intensive rotational systems, where paddock-level PB can decline by 500 to 1,500 kg DM/ha within short timeframes. It follows that this limitation might be overcome as new and upgraded satellites are deployed in the future. Supervised classification models experienced 24.2% average performance degradation when transferred across production years, highlighting sensitivity to interannual variability

in weather patterns and management practices. Furthermore, systematic measurement bias in rising plate meter calibrations propagates directly through multiquadric interpolation workflows, meaning that any inherent errors in ground truth measurements become embedded within satellite-derived estimates and require correction strategies going forward (Murphy et al., 2022).

Finally, smartphone-based PB estimation in Chapter 7 demonstrated severe spatial generalisation failure that limits immediate scalability. Leave-one-farm-out cross-validation revealed mean R^2 of -0.009 compared to within-dataset test performance of $R^2 = 0.561$, exposing farm-specific dependencies that prevent application to new farms without additional training. This spatial limitation was compounded by several data collection constraints. Single smartphone images captured per paddock cannot represent spatial heterogeneity arising from soil variation, topographic gradients, and differential grazing pressure. Additionally, the absence of point-level GPS coordinates for individual rising plate meter measurements prevented linking ground truth to specific locations within paddocks, constraining model training to paddock-averaged values rather than exploiting fine-scale spatial patterns. The training dataset of 3,291 images, whilst unique (15 farms and paddocks within them systematically sampled and monitored over 2 years) and sufficient for transfer learning with frozen features, proved inadequate for training deep learning (DL) architectures from scratch, suggesting that substantially larger datasets spanning diverse farm environments may be required to improve spatial generalisation. Although this was an undesirable outcome, the results constitute a substantial contribution to science. Additionally, rising plate meter ground truth measurements contain inherent uncertainty, with calibration relationships typically explaining 70% to 80% of PB variance (Earle & McGowan, 1979), establishing a fundamental ceiling on achievable prediction accuracy regardless of modelling sophistication. Furthermore, the scale of data collection, model development, and validation required for robust spatial generalisation exceeds individual farm capacity, necessitating institutional support from research organisations, universities, and government agencies that individual farmers cannot provide independently.

Beyond these technical and institutional constraints, the practical adoption of integrated precision management systems in commercial pasture-based dairy operations raises broader considerations that warrant acknowledgement. Reliance on satellite imagery, sensor data, and machine learning models for feed budgeting and grazing management decisions introduces a form of technological dependency that did not exist under traditional management practices, where these decisions were grounded in direct observation and accumulated farm-level experience. Over time, the progressive transfer of these decisions to data-driven systems risks eroding the practical knowledge that experienced dairy producers develop through sustained engagement with their herds and pastures. Furthermore, the data infrastructure, technical expertise, and ongoing investment required for fully integrated systems may favour larger commercial operations, concentrating the productivity advantages of precision

management among farms with greater resources, whilst smaller pasture-based operations may lack the capacity to fully participate.

FUTURE RESEARCH DIRECTIONS

Pasture-based dairy farming stands at a turning point where AI can transform how farmers manage their operations, moving from reactive decision-making to predictive systems that optimise every paddock and every cow. The integrated precision management framework illustrated in Fig. 1.1 shows what a dairy farm with fully integrated precision management could look like, where advanced monitoring systems can accurately estimate PB, track growth rates, quantify availability, and measure utilisation across every paddock and farm. Automated detection systems identify when and where each herd grazes, how much pasture was harvested, and what the actual intake patterns look like throughout the day. This information combined with environmental data and individual cow characteristics to inform precise resource allocation decisions, determining optimal concentrate supplementation for each animal based on actual pasture intake and production potential. Imagine the power this information would give farmers, knowing exactly which paddocks have sufficient feed, understanding why cows might not be eating in certain areas, identifying underperforming pastures before they become problems, and adjusting feeding strategies in real-time rather than reacting days later when milk production has already dropped. This could represent a fundamental shift from labour-intensive manual monitoring and flat-rate feeding toward data-driven management that maximises milk production whilst minimising feed waste and labour costs. However, realising this vision requires addressing the aforementioned fundamental limitations identified through five strategic research priorities that would enable the closed-loop management system necessary for transforming pasture-based dairy production.

First, multi-sensor fusion achieving daily temporal resolution represents the most critical technical advancement required for operational grazing detection and utilisation monitoring. Future research must prioritise integration of synthetic aperture radar with optical imagery to overcome cloud interference and insufficient revisit frequency that currently exclude a lot of grazing events from validation. Synthetic aperture radar systems provide all-weather capabilities independent of cloud cover, whilst optical sensors capture detailed vegetation spectral characteristics, and combining these complementary data sources enables continuous monitoring regardless of atmospheric conditions (Holtgrave et al., 2023; Lobert et al., 2021; Mercier et al., 2019). Whilst current freely available platforms like Sentinel-1 and Sentinel-2 demonstrate this potential, future satellite constellations and emerging high-frequency systems may offer enhanced temporal coverage and spatial resolution. However, commercial viability requires that these systems remain cost-effective for farmers, either through continued open-access data policies or affordable subscription models that deliver value exceeding traditional monitoring costs. Development efforts should focus on cross-sensor calibration algorithms enabling consistent PB

estimation across different radar and optical measurement systems, and temporal interpolation methods that preserve abrupt PB removal signals that occur during grazing events rather than smoothing them through averaging. Dense temporal coverage would facilitate precise identification of grazing events through continuous PB trajectory monitoring whilst enabling discrimination between gradual senescence and rapid pasture harvest patterns.

Second, automated ground truth collection infrastructure would provide a transformative opportunity eliminating labour-intensive manual measurements that constrain satellite estimation accuracy. Machine learning models are only as accurate as the ground truth data used to train them, and manual rising plate meter measurements suffer from operator variability, spatial sampling inconsistencies, and temporal gaps that limit model performance. Future research should develop and validate autonomous robotic platforms equipped with multiple sensors including rising plate meters, ultrasonic sensors, and spectral measurement devices capable of collecting georeferenced measurements across paddocks on regular schedules with minimal human intervention, reducing labour requirements to periodic calibration whilst automating routine data collection and providing the precision and consistency essential for training robust estimation models (Correa-Luna et al., 2024). These robotic systems equipped with sensors could traverse paddocks systematically, recording measurements with precise geolocation enabling direct linkage to satellite pixels whilst generating dense spatial and temporal datasets enriching training data. Because environmental conditions change continuously throughout the production cycle, automated ML pipelines ingesting robotic measurements and retraining estimation models would enable near real-time updates maintaining accuracy as conditions evolve, addressing temporal drift when models encounter novel environmental scenarios (Azubuike et al., 2025b).

Third, controlled experimental validation establishing causal evidence for precision feeding optimisation demands systematic field trials implementing machine learning-optimised recommendations through automated feeding systems compared against flat-rate controls. Future research must evaluate alternative multi-objective evolutionary algorithms beyond NSGA-II and SPEA-II tested in this thesis, incorporating broader cow characteristics including body condition score, genetic merit, walking distance, track conditions, and behavioural sensor data to enhance model robustness (Azubuike et al., 2025a). Critically, trials require substantially more training data per individual cow enabling models to learn effectively from historical patterns, whilst systematic concentrate manipulation across sufficient animals, duration, and farm diversity would definitively establish whether productivity gains materialise under active experimental conditions. Development of explainable AI methodologies providing interpretable justifications for feeding allocation decisions would address farmer concerns about algorithmic transparency, demonstrating how individual decisions aggregate to herd-level outcomes.

Fourth, advanced DL architectures and substantially expanded training datasets offer essential pathways for improving spatial transferability and prediction accuracy. Future research should prioritise advanced time-series models including transformers, autoencoders, and recurrent architectures with LSTM layers that exploit temporal image sequences to predict PB trajectories, attention mechanisms automatically identifying PB-relevant image regions, and Bayesian DL providing uncertainty quantification through prediction confidence intervals supporting risk-based decision-making. Three-dimensional imaging approaches using stereo cameras or structure-from-motion photogrammetry may improve estimation for prostrate species by capturing canopy height information unavailable from RGB imagery. Expanding training datasets to thousands of images spanning diverse farm environments, soil types, pasture species compositions, and management practices may enable end-to-end fine-tuning of DL architectures trained from scratch, improving spatial generalisation beyond current transfer learning approaches with frozen features. Progressive training strategies continuously updating model parameters with site-specific observations would provide data-efficient adaptation pathways, requiring development of standardised protocols establishing minimum ground truth requirements whilst minimising farmer time investments.

As progress is made toward addressing these barriers, integration into unified precision management platforms depicted in Fig. 1.1 becomes feasible. Future research must develop software architectures seamlessly integrating validated components within farmer-accessible interfaces translating complex spatial datasets into actionable management recommendations. Integration frameworks should handle heterogeneous data streams from satellites, smartphones, automated feeding systems, and behavioural sensors, whilst incorporating economic optimisation algorithms demonstrating value proposition across diverse farm scales. The pathway forward requires sustained research investment, industry collaboration, and farmer engagement, but the demonstrated capabilities and clear priorities established through this work provide confidence that artificial intelligence has potential to transform pasture-based dairy farming, delivering productivity gains essential for addressing the efficiency and sustainability challenges facing dairy systems globally.

REFERENCES

Akintan, O. A., Gebremedhin, K. G., & Uyeh, D. D. (2025). Linking Animal Feed Formulation to Milk Quantity, Quality, and Animal Health Through Data-Driven Decision-Making. *Animals*, *15*(2), 162. https://doi.org/10.3390/ani15020162

Ali, J., Haoran, W., Mehmood, K., Hussain, W., Iftikhar, F., Shahzad, F., Hussain, K., Qun, Y., & Zhongkui, J. (2025). Remote sensing and integration of machine learning algorithms for above-ground biomass estimation in Larix principis-rupprechtii Mayr plantations: a case study using Sentinel-2 and Landsat-9 data in northern China [Original Research]. *Frontiers in Environmental Science*, *13*. https://doi.org/10.3389/fenvs.2025.1577298

Alvarez-Mendoza, C. I., Guzman, D., Casas, J., Bastidas, M., Polanco, J., Valencia-Ortiz, M., Montenegro, F., Arango, J., Ishitani, M., & Selvaraj, M. G. (2022). Predictive Modeling of Above-Ground Biomass in Brachiaria Pastures fro m Satellite and UAV Imagery Using Machine Learning Approaches. *Remote Sensing*, *14*(22), 5870. https://doi.org/10.3390/rs14225870

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, *10*(1), 46. https://doi.org/10.1186/s40537-023-00727-2

Anderson, D. (2007). Virtual Fencing—Past, Present and Future. *Rangeland Journal*, *29*. https://doi.org/10.1071/RJ06036

André, G., Berentsen, P. B. M., Van Duinkerken, G., Engel, B., & Lansink, A. G. J. M. O. (2010). Economic potential of individual variation in milk yield response to concentrate intake of dairy cows. *The Journal of Agricultural Science*, *148*(3), 263-276. https://doi.org/10.1017/S0021859610000134

Aquilani, C., Confessore, A., Bozzi, R., Sirtori, F., & Pugliese, C. (2022). Review: Precision Livestock Farming technologies in pasture-based livestock systems. *Animal*, *16*(1), 100429. https://doi.org/10.1016/j.animal.2021.100429

Askar, Nuthammachot, N., Phairuang, W., Wicaksono, P., & Sayektiningsih, T. (2018). Estimating Aboveground Biomass on Private Forest Using Sentinel-2 Imagery. *Journal of Sensors*, *2018*(1), 6745629. https://doi.org/10.1155/2018/6745629

Askari, M. S., McCarthy, T., Magee, A., & Murphy, D. J. (2019). Evaluation of Grass Quality under Different Soil Management Scenarios Using Remote Sensing Techniques. *Remote Sensing*, *11*(15), 1835. https://doi.org/10.3390/rs11151835

Azubuike, B. N., Chlingaryan, A., Correa-Luna, M., Clark, C. E. F., & Garcia, S. C. (2025a). A Data-Driven Approach for Optimising Supplement Allocation to Individual Lactating Dairy Cows

in Pasture-Based Systems. *Smart Agricultural Technology*, 101669. https://doi.org/10.1016/j.atech.2025.101669

Azubuike, B. N., Chlingaryan, A., Correa-Luna, M., Clark, C. E. F., & Garcia, S. C. (2025b). Data Augmentation and Interpolation Improves Machine Learning-Based Pasture Biomass Estimation from Sentinel-2 Imagery. *Remote Sensing*, *17*(23), 3787. https://doi.org/10.3390/rs17233787

Bazzo, C. O. G., Kamali, B., Hütt, C., Bareth, G., & Gaiser, T. (2023). A Review of Estimation Methods for Aboveground Biomass in Grasslands Using UAV. *Remote Sensing*, *15*(3), 639. https://doi.org/10.3390/rs15030639

Bellingeri, A., Gallo, A., Liang, D., Masoero, F., & Cabrera, V. E. (2020). Development of a linear programming model for the optimal allocation of nutritional resources in a dairy herd. *J Dairy Sci*, *103*(11), 10898-10916. https://doi.org/10.3168/jds.2020-18157

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 79-87. https://doi.org/10.1016/j.jag.2015.02.012

Beukes, P., McCarthy, S., Wims, C., Gregorini, P., & Romera, A. (2018). Regular estimates of herbage mass can improve profitability of pasture-based dairy systems. *Animal Production Science*, *59*, 359-367. https://doi.org/10.1071/AN17166

Brito, L. F., Heringstad, B. r., Klaas, I. C., Schodl, K., Cabrera, V. E., Stygar, A., Iwersen, M., Haskell, M. J., Stock, K. F., Gengler, N., Bewley, J., Hostens, M., Vasseur, E., & Egger-Danner, C. (2025). Invited Review: Using data from sensors and other precision farming technologies to enhance the sustainability of dairy cattle breeding programs. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2025-26554

Britt, J. H., Cushman, R. A., Dechow, C. D., Dobson, H., Humblot, P., Hutjens, M. F., Jones, G. A., Ruegg, P. S., Sheldon, I. M., & Stevenson, J. S. (2018). Invited review: Learning from the future—A vision for dairy farms and cows in 2067. *Journal of Dairy Science*, *101*(5), 3722-3741. https://doi.org/10.3168/jds.2017-14025

Cabrera, V. E. (2025). Artificial intelligence applied to dairy science: insights from the Dairy Brain Initiative. *Animal Frontiers*, *14*(6), 60-63. https://doi.org/10.1093/af/vfae040

Cabrera, V. E., & Fadul-Pacheco, L. (2021). Future of dairy farming from the Dairy Brain perspective: Data integration, analytics, and applications. *International Dairy Journal*, *121*, 105069. https://doi.org/10.1016/j.idairyj.2021.105069

Cai, T., Chang, C., Zhao, Y., Wang, X., Yang, J., Dou, P., Otgonbayar, M., Zhang, G., Zeng, Y., & Wang, J. (2024). Within-season estimates of 10 m aboveground biomass based on Landsat, Sentinel-2 and PlanetScope data. *Scientific Data*, *11*(1), 1276. https://doi.org/10.1038/s41597-024-04120-3

Campos, L. M., Ringer, H., Chung, M., & Hanigan, M. D. (2023). Application of a mathematical framework for the optimization of precision-fed dairy cattle diets. *Animal*, *17*, 101001. https://doi.org/10.1016/j.animal.2023.101001

Cândido, B., Mindala, U., Ebrahimy, H., Zhang, Z., & Kallenbach, R. (2025). Integrating Proximal and Remote Sensing with Machine Learning for Pasture Biomass Estimation. *Sensors*, *25*(7), 1987. https://www.mdpi.com/1424-8220/25/7/1987

Cao, L., Xu, L., & Goodman, E. D. (2016). A Guiding Evolutionary Algorithm with Greedy Strategy for Global Optimization Problems. *Comput Intell Neurosci*, *2016*, 2565809. https://doi.org/10.1155/2016/2565809

Capolupo, A., Kooistra, L., Berendonk, C., Boccia, L., & Suomalainen, J. (2015). Estimating Plant Traits of Grasslands from UAV-Acquired Hyperspectral Images: A Comparison of Statistical Approaches. *ISPRS International Journal of Geo-Information*, *4*(4), 2792-2820. https://doi.org/10.3390/ijgi4042792

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, *62*(3), 241-252. https://doi.org/10.1016/S0034-4257(97)00104-1

Castro, W., Marcato Junior, J., Polidoro, C., Osco, L. P., Gonçalves, W., Rodrigues, L., Santos, M., Jank, L., Barrios, S., Valle, C., Simeão, R., Carromeu, C., Silveira, E., Jorge, L. A. d. C., & Matsubara, E. (2020). Deep Learning Applied to Phenotyping of Biomass in Forages with UAV-Based RGB Imagery. *Sensors*, *20*(17), 4802. https://doi.org/10.3390/s20174802

Cavallini, D., Giammarco, M., Buonaiuto, G., Vignola, G., De Matos Vettori, J., Lamanna, M., Prasinou, P., Colleluori, R., Formigoni, A., & Fusaro, I. (2025). Two years of precision livestock management: harnessing ear tag device behavioral data for pregnancy detection in free-range dairy cattle on silage/hay-mix ration [Original Research]. *Frontiers in Animal Science*, *Volume 6 - 2025*. https://doi.org/10.3389/fanim.2025.1547395

Chapman, D., Bryant, J., Olayemi, M., Edwards, G., Thorrold, B., McMillan, W., Kerr, G., Judson, G., Cookson, T., Moorhead, A., & Norriss, M. (2016). An economically based evaluation index for perennial and short-term ryegrasses in New Zealand dairy farm systems. *Grass and Forage Science*, *72*, n/a-n/a. https://doi.org/10.1111/gfs.12213

Chebli, Y., El Otmani, S., Hornick, J.-L., Keli, A., Bindelle, J., Chentouf, M., & Cabaraux, J.-F. (2022). Using GPS Collars and Sensors to Investigate the Grazing Behavior and Energy Balance of Goats Browsing in a Mediterranean Forest Rangeland. *Sensors*, *22*(3), 781. https://doi.org/10.3390/s22030781

Chen, S., Wang, J., Liu, Q., Liang, X., Liu, R., Qin, P., Yuan, J., Wei, J., Yuan, S., Huang, H., & Gong, P. (2024). Global 30 m seamless data cube (2000–2022) of land surface reflectance generated from Landsat 5, 7, 8, and 9 and MODIS Terra constellations. *Earth Syst. Sci. Data*, *16*(11), 5449-5475. https://doi.org/10.5194/essd-16-5449-2024

Chen, Y., Guerschman, J., Shendryk, Y., Henry, D., & Harrison, M. T. (2021). Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learni ng. *Remote Sensing*, *13*(4), 603. https://doi.org/10.3390/rs13040603

Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, *9*(4), 414-424. https://doi.org/10.1016/j.jag.2007.02.001

Clark, C. E. F., Kaur, R., Millapan, L. O., Golder, H. M., Thomson, P. C., Horadagoda, A., Islam, M. R., Kerrisk, K. L., & Garcia, S. C. (2018). The effect of temperate or tropical pasture grazing state and grain-based concentrate allocation on dairy cattle production and behavior. *Journal of Dairy Science*, *101*(6), 5454-5465. https://doi.org/10.3168/jds.2017-13388

Cockburn, M. (2020). Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms. *Animals*, *10*(9), 1690. https://www.mdpi.com/2076-2615/10/9/1690

Correa-Luna, M., Gargiulo, J., Beale, P., Deane, D., Leonard, J., Hack, J., Geldof, Z., Wilson, C., & Garcia, S. (2024). Accounting for minimum data required to train a machine learning model to accurately monitor Australian dairy pastures using remote sensing. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-68094-3

Craig, A.-L., Gordon, A. W., Hamill, G., & Ferris, C. P. (2022). Milk Composition and Production Efficiency within Feed-To-Yield Systems on Commercial Dairy Farms in Northern Ireland. *Animals*, *12*(14), 1771. https://doi.org/10.3390/ani12141771

Das, R., Nath Das, K., & Mallik, S. (2023). An efficient evolutionary optimizer for solving complex dairy feed optimization problems. *Computers and Electronics in Agriculture*, *204*, 107566. https://doi.org/10.1016/j.compag.2022.107566

Davison, C., Bowen, J. M., Michie, C., Rooke, J. A., Jonsson, N., Andonovic, I., Tachtatzis, C., Gilroy, M., & Duthie, C. A. (2021). Predicting feed intake using modelling based on feeding behaviour in finishing beef steers. *Animal*, *15*(7), 100231. https://doi.org/10.1016/j.animal.2021.100231

De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P., & Rowlings, D. (2021). Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, *180*. https://doi.org/10.1016/j.compag.2020.105880

De Vries, A., Bliznyuk, N., & Pinedo, P. (2023). Invited Review: Examples and opportunities for artificial intelligence (AI) in dairy farms*. *Applied Animal Science*, *39*(1), 14-22. https://doi.org/10.15232/aas.2022-02345

Dela Rue, B. T., & Eastwood, C. R. (2017). Individualised feeding of concentrate supplement in pasture-based dairy systems: practices and perceptions of New Zealand dairy farmers and

their advisors. *Animal Production Science*, *57*(7), 1543-1549. https://doi.org/10.1071/AN16471

Distante, D., Albanello, C., Zaffar, H., Faralli, S., & Amalfitano, D. (2025). Artificial intelligence applied to precision livestock farming: A tertiary study. *Smart Agricultural Technology*, *11*, 100889. https://doi.org/10.1016/j.atech.2025.100889

Dlamini, C. M., Odindi, J., Matongera, T. N., & Mutanga, O. (2025). Exploring the utility of remote sensing technology in vegetation below ground biomass (BGB) estimation: a critical review of methods and challenges [Review]. *Frontiers in Remote Sensing*, *Volume 6 - 2025*. https://doi.org/10.3389/frsen.2025.1668676

Dos Reis, A. A., Werner, J. P. S., Silva, B. C., Figueiredo, G. K. D. A., Antunes, J. F. G., Esquerdo, J. C. D. M., Coutinho, A. C., Lamparelli, R. A. C., Rocha, J. V., & Magalhães, P. S. G. (2020). Monitoring Pasture Aboveground Biomass and Canopy Height in an Integrated Crop–Livestock System Using Textural Information from PlanetScope Imagery. *Remote Sensing*, *12*(16), 2534. https://www.mdpi.com/2072-4292/12/16/2534

Dowson, O., Philpott, A., Mason, A., & Downward, A. (2019). A multi-stage stochastic optimization model of a pastoral dairy farm. *European Journal of Operational Research*, *274*(3), 1077-1089. https://doi.org/10.1016/j.ejor.2018.10.033

Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing [Specialty Grand Challenge]. *Frontiers in Remote Sensing*, *Volume 2 - 2021*. https://doi.org/10.3389/frsen.2021.619818

Earle, D. F., & McGowan, A. A. (1979). Evaluation and calibration of an automated rising plate meter for estimating dry matter yield of pasture. *Australian Journal of Experimental Agriculture*, *19*, 337-343. https://doi.org/10.1071/EA9790337

Fernandes, M., FernandesJunior, J. S., Adams, J. M., Lee, M., Reis, R. A., & Tedeschi, L. O. (2024). Using sentinel-2 satellite images and machine learning algorithms to predict tropical pasture forage mass, crude protein, and fiber content. *Sci Rep*, *14*(1), 8704. https://doi.org/10.1038/s41598-024-59160-x

Filippelli, S., Schleeweis, K., Nelson, M., Fekety, P., & Vogeler, J. (2024). Testing temporal transferability of remote sensing models for large area monitoring. *Science of Remote Sensing*, *9*, 100119. https://doi.org/10.1016/j.srs.2024.100119

Franzoi, R. E., Menezes, B. C., Kelly, J. D., & Swartz, C. L. E. (2022). Adaptive least-squares surrogate modeling for reaction systems. In Y. Yamashita & M. Kano (Eds.), *Computer Aided Chemical Engineering* (Vol. 49, pp. 1705-1710). Elsevier. https://doi.org/10.1016/B978-0-323-85159-6.50284-0

Freitas, R. G., Pereira, F. R. S., Dos Reis, A. A., Magalhães, P. S. G., Figueiredo, G. K. D. A., & do Amaral, L. R. (2022). Estimating pasture aboveground biomass under an integrated crop-livestock system based on spectral and texture measures derived from UAV images.

Computers and Electronics in Agriculture, *198*, 107122. https://doi.org/10.1016/j.compag.2022.107122

Gaffney, R., Porensky, L. M., Gao, F., Irisarri, J. G., Durante, M., Derner, J. D., & Augustine, D. J. (2018). Using APAR to Predict Aboveground Plant Productivity in Semi-Arid Rangelands: Spatial and Temporal Relationships Differ. *Remote Sensing*, *10*(9).

García, R., Aguilar, J., Pinto, A., & Toro, M. (2020). A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, *179*. https://doi.org/10.1016/j.compag.2020.105826

García, S., Islam, M., Clark, C., & Martin, P. (2014). Kikuyu-based pasture for dairy production: a review. *Crop and Pasture Science*, *65*(8), 787-797. https://doi.org/10.1071/CP13414

García, S. C., Pedernera, M., Fulkerson, W. J., Horadagoda, A., & Nandra, K. (2007). Feeding concentrates based on individual cow requirements improves the yield of milk solids in dairy cows grazing restricted pasture. *Australian Journal of Experimental Agriculture*, *47*(5), 502-508. https://doi.org/10.1071/EA05349

Gard, S., Neal, M., & Minnee, E. (2024). Pasture performance tools: Current and future state. *Journal of New Zealand Grasslands*, *86*, 273-279. https://doi.org/10.33584/jnzg.2024.86.3688

Gargiulo, J., Clark, C., Lyons, N., de Veyrac, G., Beale, P., & Garcia, S. (2020). Spatial and Temporal Pasture Biomass Estimation Integrating Electronic Plate Meter, Planet CubeSats and Sentinel-2 Satellite Data. *Remote Sensing*, *12*(19), 3222. https://doi.org/10.3390/rs12193222

Gargiulo, J. I., Lyons, N. A., Masia, F., Beale, P., Insua, J. R., Correa-Luna, M., & Garcia, S. C. (2023). Comparison of Ground-Based, Unmanned Aerial Vehicles and Satellite Rem ote Sensing Technologies for Monitoring Pasture Biomass on Dairy Farms. *Remote Sensing*, *15*(11), 2752. https://doi.org/10.3390/rs15112752

Greenwood, P. L., Gardner, G. E., & Ferguson, D. M. (2018). Current situation and future prospects for the Australian beef industry - A review. *Asian-Australas J Anim Sci*, *31*(7), 992-1006. https://doi.org/10.5713/ajas.18.0090

He, K., Zhang, X., Ren, S., & Sun, J. (2016, 27-30 June 2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Heins, B. J., Pereira, G. M., & Sharpe, K. T. (2023). Precision technologies to improve dairy grazing systems* *Presented as part of the Joint ADSA Midwest Branch/Forages and Pastures Symposium: Grazing to Improve Profitability of Midwest Dairy Farms held at the ADSA Annual Meeting, June 2022. *JDS Communications*, *4*(4), 318-323. https://doi.org/10.3168/jdsc.2022-0308

Herlin, A., Brunberg, E., Hultgren, J., Högberg, N., Rydberg, A., & Skarin, A. (2021). Animal Welfare Implications of Digital Tools for Monitoring and Management of Cattle and Sheep on Pasture. *Animals*, *11*(3), 829. https://www.mdpi.com/2076-2615/11/3/829

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., & Bonfil, D. J. (2011). LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands. *Remote Sensing of Environment*, *115*(8), 2141-2151. https://doi.org/10.1016/j.rse.2011.04.018

Hills, J. L., García, S. C., Dela Rue, B., & Clark, C. E. F. (2015). Limitations and potential for individualised feeding of concentrate supplements to grazing dairy cows. *Animal Production Science*, *55*(7), 922-930. https://doi.org/10.1071/AN14855

Holtgrave, A.-K., Lobert, F., Erasmi, S., Röder, N., & Kleinschmit, B. (2023). Grassland mowing event detection using combined optical, SAR, and weather time series. *Remote Sensing of Environment*, *295*, 113680. https://doi.org/10.1016/j.rse.2023.113680

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695. https://doi.org/10.1007/s12525-021-00475-2

Ji, W., Luo, Y., Liao, Y., Wu, W., Wei, X., Yang, Y., He, X. Z., Shen, Y., Ma, Q., Yi, S., & Sun, Y. (2023). UAV Assisted Livestock Distribution Monitoring and Quantification: A Low-Cost and High-Precision Solution. *Animals*, *13*(19), 3069. https://doi.org/10.3390/ani13193069

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, *155*, 269-288. https://doi.org/10.1016/j.agsy.2016.09.021

Kallenbach, R., Hamilton, S., & Lock, T. R. (2020). 345 Digital Technologies Ease the Burden of Pasture Intake Measurements: A New Hope. *Journal of Animal Science*, *98*(Supplement_4), 78-78. https://doi.org/10.1093/jas/skaa278.142

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

Kamphuis, C., Riel, J. W. v., Veerkamp, R. F., & Mol, R. M. d. (2017). Traditional mixed linear modelling versus modern machine learning to estimate cow individual feed intake. Precision Livestock Farming '17,

Kaur, U., Malacco, V. M. R., Bai, H., Price, T. P., Datta, A., Xin, L., Sen, S., Nawrocki, R. A., Chiu, G., Sundaram, S., Min, B.-C., Daniels, K. M., White, R. R., Donkin, S. S., Brito, L. F., & Voyles, R. M. (2023). Invited review: integration of technologies and systems for precision animal agriculture—a case study on precision dairy farming. *Journal of Animal Science*, *101*. https://doi.org/10.1093/jas/skad206

Kleen, J. L., & Guatteo, R. (2023). Precision Livestock Farming: What Does It Contain and What Are the Perspectives? *Animals*, *13*(5), 779. https://doi.org/10.3390/ani13050779

Kolver, E. S. (2003). Nutritional limitations to increased production on pasture-based systems. *Proceedings of the Nutrition Society*, *62*(2), 291-300. https://doi.org/10.1079/PNS2002200

Krizsan, S. J., Sairanen, A., Höjer, A., & Huhtanen, P. (2014). Evaluation of different feed intake models for dairy cows. *J Dairy Sci*, *97*(4), 2387-2397. https://doi.org/10.3168/jds.2013-7561

Lamanna, M., Bovo, M., Bellisola, G., Romanzin, A., & Cavallini, D. (2025). Rethinking Wearable Technology in Dairy Cows: Challenges and Prospects for Smart Collars. *The Open Agriculture Journal*, *19*. https://doi.org/10.2174/0118743315410860250914045935

Lamanna, M., Bovo, M., & Cavallini, D. (2025). Wearable Collar Technologies for Dairy Cows: A Systematized Review of the Current Applications and Future Innovations in Precision Livestock Farming. *Animals*, *15*(3), 458. https://doi.org/10.3390/ani15030458

Laplacette, C. M., Berone, G. D., Utsumi, S. A., & Insua, J. R. (2025). Calibration of an Unmanned Aerial Vehicle for Prediction of Herbage Mass in Temperate Pasture. *Agriculture*, *15*(5), 492. https://doi.org/10.3390/agriculture15050492

Lawson, A. R., Giri, K., Thomson, A. L., Karunaratne, S. B., Smith, K. F., Jacobs, J. L., & Morse-McNabb, E. M. (2022). Multi-site calibration and validation of a wide-angle ultrasonic senso r and precise GPS to estimate pasture mass at the paddock scale. *Computers and Electronics in Agriculture*, *195*, 106786. https://doi.org/10.1016/j.compag.2022.106786

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Li, J., Kebreab, E., You, F., Fadel, J. G., Hansen, T. L., VanKerkhove, C., & Reed, K. F. (2022). The application of nonlinear programming on ration formulation for dairy cattle [Article]. *Journal of Dairy Science*, *105*(3), 2180-2189. https://doi.org/10.3168/jds.2021-2081710.3168/jds.2021-20817

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, *18*(8), 2674. https://www.mdpi.com/1424-8220/18/8/2674

Liebe, D. M., & White, R. R. (2019). Analytics in sustainable precision animal nutrition. *Animal Frontiers: The Review Magazine of Animal Agriculture*, *9*(2), 16 - 24. https://doi.org/10.1093/af/vfz003

Little, M. W., O'Connell, N. E., & Ferris, C. P. (2016). A comparison of individual cow versus group concentrate allocation strategies on dry matter intake, milk production, tissue changes, and fertility of Holstein-Friesian cows offered a grass silage diet. *J Dairy Sci*, *99*(6), 4360-4373. https://doi.org/10.3168/jds.2015-10441

Lobert, F., Holtgrave, A.-K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., & Erasmi, S. (2021). Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. *Remote Sensing of Environment*, *267*, 112751. https://doi.org/10.1016/j.rse.2021.112751

Lokhorst, C., de Mol, R. M., & Kamphuis, C. (2019). Invited review: Big Data in precision dairy farming. *Animal*, *13*(7), 1519-1528. https://doi.org/10.1017/S1751731118003439

Lussem, U., Bolten, A., Menne, J., Gnyp, M., Schellberg, J., & Bareth, G. (2019). Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *Journal of Applied Remote Sensing*, *13*, 1. https://doi.org/10.1117/1.JRS.13.034525

Lussem, U., Schellberg, J., & Bareth, G. (2020). Monitoring Forage Mass with Low-Cost UAV Data: Case Study at the Rengen Grassland Experiment. *PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science*, *88*. https://doi.org/10.1007/s41064-020-00117-w

Mahato, S., & Neethirajan, S. (2024). Integrating Artificial Intelligence in dairy farm management − biometric facial recognition for cows. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2024.10.001

Martin, M. J., Dórea, J. R. R., Borchers, M. R., Wallace, R. L., Bertics, S. J., DeNise, S. K., Weigel, K. A., & White, H. M. (2021). Comparison of methods to predict feed intake and residual feed intake using behavioral and metabolite data in addition to classical performance variables. *Journal of Dairy Science*, *104*(8), 8765-8782. https://doi.org/10.3168/jds.2020-20051

McCarthy, A., Raedts, P., Foley, J., & Hills, J. (2022, Australia 30 Nov - 02 Dec 2022). *Improving pasture growth assessment using machine vision* Australasian Dairy Science Symposium 2022 (ADSS 2022), Twin Waters, Australia. https://research.usq.edu.au/item/q7x3x/improving-pasture-growth-assessment-using-machine-vision

Melak, A., Aseged, T., & Shitaw, T. (2024). The Influence of Artificial Intelligence Technology on the Management of Livestock Farms. *International Journal of Distributed Sensor Networks*, *2024*(1), 8929748. https://doi.org/10.1155/2024/8929748

Mercier, A., Betbeder, J., Rumiano, F., Baudry, J., Gond, V., Blanc, L., Bourgoin, C., Cornu, G., Ciudad, C., Marchamalo, M., Poccard-Chapuis, R., & Hubert-Moy, L. (2019). Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest–Agriculture Mosaics in Temperate and Tropical Landscapes. *Remote Sensing*, *11*(8), 979.

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620-1633. https://doi.org/10.1111/2041-210X.13650

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*, 1-9. https://doi.org/10.1016/j.envsoft.2017.12.001

Mijić, D., Vico, G., Popović, B., Popović, N., Ljubojević, M., & Savić, M. (2024). OPTIMILK: A Web-Based Tool for Least-Cost Dairy Ration Optimization Using Linear Programming. *Agriculture*, *14*(9), 1580. https://doi.org/10.3390/agriculture14091580

Mike, Z., Juan, A. C., Diana, M. G.-Z., Mario, C.-M., Jhon, F. G., Christoph, R., Nicholas, J., Miller, E., Kenny, R., & Brian, B. (2024). Pixels to Pasture: Using Machine Learning and Multispectral Remote Sen sing to Predict Biomass and Nutrient Quality in Tropical Grasslands. *Remote Sensing Applications Society and Environment*. https://doi.org/10.1016/j.rsase.2024.101282

Monteiro, H. F., Figueiredo, C. C., Mion, B., Santos, J. E. P., Bisinotto, R. S., Peñagaricano, F., Ribeiro, E. S., Marinho, M. N., Zimpel, R., da Silva, A. C., Oyebade, A., Lobo, R. R., Coelho Jr, W. M., Peixoto, P. M. G., Ugarte Marin, M. B., Umaña-Sedó, S. G., Rojas, T. D. G., Elvir-Hernandez, M., Schenkel, F. S., . . . Lima, F. S. (2024). An artificial intelligence approach of feature engineering and ensemble methods depicts the rumen microbiome contribution to feed efficiency in dairy cows. *Animal Microbiome*, *6*(1), 5. https://doi.org/10.1186/s42523-024-00289-5

Morrone, S., Dimauro, C., Gambella, F., & Cappai, M. G. (2022). Industry 4.0 and Precision Livestock Farming (PLF): An up to Date Overview across Animal Productions. *Sensors*, *22*(12), 4319. https://doi.org/10.3390/s22124319

Morse-McNabb, E. M., Hasan, M. F., & Karunaratne, S. (2023). A Multi-Variable Sentinel-2 Random Forest Machine Learning Model Appro ach to Predicting Perennial Ryegrass Biomass in Commercial Dairy Farms in Southeast Australia. *Remote Sensing*, *15*(11), 2915. https://doi.org/10.3390/rs15112915

Moscovici Joubran, A., Pierce, K. M., Garvey, N., Shalloo, L., & O'Callaghan, T. F. (2021). Invited review: A 2020 perspective on pasture-based dairy systems and products. *Journal of Dairy Science*, *104*(7), 7364-7382. https://doi.org/10.3168/jds.2020-19776

Murphy, D. J., O' Brien, B., O' Donovan, M., Condon, T., & Murphy, M. D. (2022). A near infrared spectroscopy calibration for the prediction of fresh grass quality on Irish pastures. *Information Processing in Agriculture*, *9*(2), 243-253. https://doi.org/10.1016/j.inpa.2021.04.012

Mutanga, O., Masenyama, A., & Sibanda, M. (2023). Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. *ISPRS Journal of Photogrammetry and Remote Sensing*, *198*, 297-309. https://doi.org/10.1016/j.isprsjprs.2023.03.010

Narayanan, B., Saadeldin, M., Albert, P., McGuinness, K., & Mac Namee, B. (2021). Extracting pasture phenotype and biomass percentages using weakly supervised multi-target deep learning on a small dataset. *arXiv preprint arXiv:2101.03198*. https://doi.org/10.48550/arXiv.2101.03198

Nasiri, V., Deljouei, A., Moradi, F., Sadeghi, S. M. M., & Borz, S. A. (2022). Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. *Remote Sensing*, *14*(9), 1977. https://doi.org/10.3390/rs14091977

Nickmilder, C., Tedde, A., Dufrasne, I., Lessire, F., Tychon, B., Curnel, Y., Bindelle, J., & Soyeurt, H. (2021). Development of Machine Learning Models to Predict Compressed Sward Hei ght in Walloon Pastures Based on Sentinel-1, Sentinel-2 and Meteorolog ical Data Using Multiple Data Transformations. *Remote Sensing*, *13*(3), 408. https://doi.org/10.3390/rs13030408

Norton, T., & Berckmans, D. (2017). Developing precision livestock farming tools for precision dairy farming. *Animal Frontiers*, *7*(1), 18-23. https://doi.org/10.2527/af.2017.0104

Notte, G., Cancela, H., Pedemonte, M., Chilibroste, P., Rossing, W. A. H., & Groot, J. C. J. (2020). A multi-objective optimization model for dairy feeding management. *Agricultural Systems*, *183*, 102854. https://doi.org/10.1016/j.agsy.2020.102854

Notte, G., Chilibroste, P., Pedemonte, M., & Cancela, H. (2021, 2-4 Nov. 2021). Evolutionary multi-objective algorithms for feed resource allocation in dairy systems. 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI),

NRC, N. R. C. (2001). *Nutrient Requirements of Dairy Cattle: Seventh Revised Edition, 2001*. The National Academies Press. https://doi.org/10.17226/9825

Numata, I., Roberts, D., Chadwick, O., Schimel, J., Galvao, L., & Soares, J. (2008). Evaluation of hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging spectrometers. *Remote Sensing of Environment*, *112*, 1569-1583. https://doi.org/10.1016/j.rse.2007.08.014

Nyamuryekung'e, S., Duff, G., Utsumi, S., Estell, R., McIntosh, M. M., Funk, M., Cox, A., Cao, H., Spiegal, S., Perea, A., & Cibils, A. F. (2023). Real-Time Monitoring of Grazing Cattle Using LORA-WAN Sensors to Improve Precision in Detecting Animal Welfare Implications via Daily Distance Walked Metrics. *Animals*, *13*(16), 2641. https://doi.org/10.3390/ani13162641

Ogungbuyi, M. G., Guerschman, J., Fischer, A. M., Crabbe, R. A., Ara, I., Mohammed, C., Scarth, P., Tickle, P., Whitehead, J., & Harrison, M. T. (2024). Improvement of pasture biomass modelling using high-resolution satelli te imagery and machine learning. *Journal of Environmental Management*, *356*, 120564. https://doi.org/10.1016/j.jenvman.2024.120564

Ogungbuyi, M. G., Guerschman, J., Fischer, A. M., Mohammed, C., Crabbe, R. A., & Harrison, M. T. (2025). Using vegetation indices from nanosatellites for timely prediction of pasture biomass. *Total Environment Advances*, *15*, 200130. https://doi.org/10.1016/j.teadva.2025.200130

Pasture.io. *Pasture.io-pasture measurement on autopilot*. Retrieved 12 December 2025 from https://pasture.io/

Peña, T., Lara, P., & Castrodeza, C. (2009). Multiobjective stochastic programming for feed formulation. *JORS*, *60*, 1738-1748. https://doi.org/10.1057/jors.2008.106

Peng, J., Zeiner, N., Parsons, D., Féret, J.-B., Söderström, M., & Morel, J. (2023). Forage Biomass Estimation Using Sentinel-2 Imagery at High Latitudes. *Remote Sensing*, *15*(9), 2350. https://doi.org/10.3390/rs15092350

Pivoto, D., Waquil, P. D., Talamini, E., Finocchio, C. P. S., Dalla Corte, V. F., & de Vargas Mores, G. (2018). Scientific development of smart farming technologies and their application in Brazil. *Information Processing in Agriculture*, *5*(1), 21-32. https://doi.org/10.1016/j.inpa.2017.12.002

Prasad, A., Peters, M., Matthews, S., & Iverson, L. (2023). Unpacking the 'black box': Improving ecological interpretation of regression-based models. *Diversity and Distributions*, *29*(7), 926-945. https://doi.org/10.1111/ddi.13707

Pullanagari, R. R., Kereszturi, G., Ian, J. Y., & Irwin, M. E. (2015, 14 October 2015). Determination of pasture quality using airborne hyperspectral imaging. Proc.SPIE,

Punalekar, S. M., Verhoef, A., Quaife, T. L., Humphries, D., Bermingham, L., & Reynolds, C. K. (2018). Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment*, *218*, 207-220. https://doi.org/10.1016/j.rse.2018.09.028

Purcell, P. J., Dale, A. J., Gordon, A. W., Barley, J., & Ferris, C. P. (2016). Effects of predicted milk yields sustained by grazed grass on dairy cow performance and concentrate requirements throughout the grazing season. *Grass and Forage Science*, *71*(3), 389-402. https://doi.org/10.1111/gfs.12193

Radeloff, V. C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., Friedl, M., Gao, F., Gorelick, N., Hansen, M., Healey, S., Hostert, P., Hulley, G., Huntington, J. L., Johnson, D. M., Neigh, C., Lyapustin, A., Lymburner, L., Pahlevan, N., . . . Zhu, Z. (2024). Need and vision for global medium-resolution Landsat and Sentinel-2 data products. *Remote Sensing of Environment*, *300*, 113918. https://doi.org/10.1016/j.rse.2023.113918

Raedts, P. J. M., & Hills, J. L. (2024). Milk yield and feeding behaviour responses to two flat-rate levels of concentrate supplementation fed over a period of 8 months to cohorts of grazing dairy cows, differing in genotype, bodyweight, or milk yield. *Animal Production Science*.

Rennie, G., King, W., Puha, M., Dalley, D. E., Dynes, R., & Upsdell, M. (2009). Calibration of the C-DAX Rapid Pasturemeter and the rising plate meter for kikuyu-based Northland dairy pastures. *Proceedings of the New Zealand Grassland Association*. https://doi.org/10.33584/jnzg.2009.71.2779

Rennie, G. M., Dalley, D. E., Dynes, R. A., & Upsdell, M. (2010). Pasture Mass Estimation by the C-DAX Pasture Meter: Regional Calibrations for New Zealand.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913-929. https://doi.org/10.1111/ecog.02881

Roche, J. R., Berry, D. P., Bryant, A. M., Burke, C. R., Butler, S. T., Dillon, P. G., Donaghy, D. J., Horan, B., Macdonald, K. A., & Macmillan, K. L. (2017). A 100-Year Review: A century of

change in temperate grazing dairy systems. *Journal of Dairy Science*, *100*(12), 10189-10233. https://doi.org/10.3168/jds.2017-13182

Russell, S., & Norvig, P. (2021). *Artificial Intelligence, Global Edition : A Modern Approach*. Pearson Deutschland. https://elibrary.pearson.de/book/99.150005/9781292401171

Rutten, C. J., Velthuis, A. G. J., Steeneveld, W., & Hogeveen, H. (2013). Invited review: Sensors to support health management on dairy farms. *Journal of Dairy Science*, *96*(4), 1928-1952. https://doi.org/10.3168/jds.2012-6107

Saar, M., Edan, Y., Godo, A., Lepar, J., Parmet, Y., & Halachmi, I. (2022). A machine vision system to predict individual cow feed intake of different feeds in a cowshed. *Animal*, *16*(1), 100432. https://doi.org/10.1016/j.animal.2021.100432

Sanderson, M., Rotz, C. A., Fultz, S., & Rayburn, E. (2001). Estimating Forage Mass with a Commercial Capacitance Meter, Rising Plate Meter, and Pasture Ruler. *Agronomy Journal*, *93*. https://doi.org/10.2134/agronj2001.1281

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Shahi, T. B., Balasubramaniam, T., Sabir, K., & Nayak, R. (2025). Pasture monitoring using remote sensing and machine learning: A review of methods and applications. *Remote Sensing Applications: Society and Environment*, *37*, 101459. https://doi.org/10.1016/j.rsase.2025.101459

Sharma, P., Leigh, L., Chang, J., Maimaitijiang, M., & Caffé, M. (2022). Above-Ground Biomass Estimation in Oats Using UAV Remote Sensing and Machine Learning. *Sensors*, *22*(2), 601. https://www.mdpi.com/1424-8220/22/2/601

Shine, P., & Murphy, M. D. (2022). Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study. *Sensors*, *22*(1), 52. https://www.mdpi.com/1424-8220/22/1/52

Shorten, P. R. (2021). Estimating milk yield for individual cows using measurements of total milk flow. *Computers and Electronics in Agriculture*, *190*(C), 106473. https://doi.org/10.1016/j.compag.2021.106473

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition* 3rd International Conference on Learning Representations (ICLR 2015), San Diego, 7-9 May 2015. arXiv:1409.1556.

Skovsen, S., Dyrmann, M., Krogh Mortensen, A., Gislum, R., Eriksen, J., Farkhani, S., Karstoft, H., & Jørgensen, R. (2019). *The GrassClover Image Dataset for Semantic and Hierarchical Species Understanding in Agriculture*. https://doi.org/10.1109/CVPRW.2019.00325

Smith, H. D., Dubeux, J. C. B., Zare, A., & Wilson, C. H. (2023). Assessing Transferability of Remote Sensing Pasture Estimates Using Multiple Machine Learning Algorithms and Evaluation Structures. *Remote Sensing*, *15*(11), 2940. https://www.mdpi.com/2072-4292/15/11/2940

Song, J.-W., Lee, M., Cho, H., Lee, D.-H., Seo, S., & Lee, W.-H. (2025). Development of individual models for predicting cow milk production for real-time monitoring. *Computers and Electronics in Agriculture*, *228*, 109698. https://doi.org/10.1016/j.compag.2024.109698

Souza, V. C., Liebe, D. M., Price, T. P., Ellett, M. D., Davis, T. C., Gleason, C. B., Daniels, K. M., & White, R. R. (2022). Algorithm development for individualized precision feeding of supplemental top dresses to influence feed efficiency of dairy cattle. *J Dairy Sci*, *105*(5), 4048-4063. https://doi.org/10.3168/jds.2021-20841

Stott, E., Williams, R. D., & Hoey, T. B. (2020). Ground Control Point Distribution for Accurate Kilometre-Scale Topographic Mapping Using an RTK-GNSS Unmanned Aerial Vehicle and SfM Photogrammetry. *Drones*, *4*(3), 55. https://doi.org/10.3390/drones4030055

Stumpe, C., Leukel, J., & Zimpel, T. (2023). Prediction of pasture yield using machine learning-based optical sensing: a systematic review. *Precision Agriculture*, *25*, 1-30. https://doi.org/10.1007/s11119-023-10079-9

Thomson, A., Jacobs, J., & Morse-McNabb, E. (2023). Comparing the predictive ability of Sentinel-2 multispectral imagery a nd a proximal hyperspectral sensor for the estimation of pasture nutri tive characteristics in an intensive rotational grazing system. *Computers and Electronics in Agriculture*, *214*, 108275. https://doi.org/10.1016/j.compag.2023.108275

Togeirode Alckmin, G., Lucieer, A., Rawnsley, R., & Kooistra, L. (2022). Perennial ryegrass biomass retrieval through multispectral UAV data. *Computers and Electronics in Agriculture*, *193*, 106574. https://doi.org/10.1016/j.compag.2021.106574

Tzanidakis, C., Tzamaloukas, O., Simitzis, P., & Panagakis, P. (2023). Precision Livestock Farming Applications (PLF) for Grazing Animals. *Agriculture*, *13*(2), 288. https://doi.org/10.3390/agriculture13020288

U, L., A, B., I, K., J, J., M, G., J, S., & G, B. (2022). Herbage Mass, N Concentration, and N Uptake of Temperate Grasslands Ca n Adequately Be Estimated from UAV-Based Image Data Using Machine Lear ning. *Remote Sensing*. https://doi.org/10.3390/rs14133066

Usigbe, M. J., Darlan, D., Uyeh, D. D., & Mallipeddi, R. (2023). *Animal Feed Optimization under Price Fluctuations using Evolutionary Algorithms* 2023 14th International Conference on Information and Communication Technology Convergence (ICTC), https://ieeexplore.ieee.org/document/10393678

Uyeh, D. D., Pamulapati, T., Mallipeddi, R., Park, T., Asem-Hiablie, S., Woo, S., Kim, J., Kim, Y., & Ha, Y. (2019). Precision animal feed formulation: An evolutionary multi-objective approach. *Animal Feed Science and Technology*, *256*, 114211. https://doi.org/10.1016/j.anifeedsci.2019.114211

Vahidi, M., Shafian, S., Thomas, S., & Maguire, R. (2023). Pasture Biomass Estimation Using Ultra-High-Resolution RGB UAVs Images and Deep Learning. *Remote Sensing*, *15*(24), 5714. https://doi.org/10.3390/rs15245714

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

Vázquez Diosdado, J. A., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., & Codling, E. A. (2015). Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, *3*(1), 15. https://doi.org/10.1186/s40317-015-0045-8

Verdon, M., Hunt, I., & Rawnsley, R. (2024). The effectiveness of a virtual fencing technology to allocate pasture and herd cows to the milking shed. *Journal of Dairy Science*, *107*(8), 6161-6177. https://doi.org/10.3168/jds.2023-24537

Verdon, M., Langworthy, A., & Rawnsley, R. (2021). Virtual fencing technology to intensively graze lactating dairy cattle. II: Effects on cow welfare and behavior. *Journal of Dairy Science*, *104*(6), 7084-7094. https://doi.org/10.3168/jds.2020-19797

Villalobos-Villalobos, L. A., & WingChing-Jones, R. (2023). Forage Biomass Estimated with a Pre-Calibrated Equation of a Rising Platemeter in Pastures Grown in Tropical Conditions. *Grasses*, *2*(2), 127-141. https://doi.org/10.3390/grasses2020011

Wang, K., Wilder, B., Perrault, A., & Tambe, M. (2020). Automatically Learning Compact Quality-aware Surrogates for Optimization Problems. *ArXiv*, *abs/2006.10815*. https://doi.org/10.48550/arXiv.2006.10815

Welchowski, T., Maloney, K. O., Mitchell, R., & Schmid, M. (2022). Techniques to Improve Ecological Interpretability of Black-Box Machine Learning Models. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(1), 175-197. https://doi.org/10.1007/s13253-021-00479-7

Williams, B., & Cremaschi, S. (2021). Novel Tool for Selecting Surrogate Modeling Techniques for Surface Approximation. In M. Türkay & R. Gani (Eds.), *Computer Aided Chemical Engineering* (Vol. 50, pp. 451-456). Elsevier. https://doi.org/10.1016/B978-0-323-88506-5.50071-1

Woodrow, L., Carter, J., Fraser, G., & Barnetson, J. (2023). Using Continuous Output Neural Nets to Estimate Pasture Biomass from Digital Photographs in Grazing Lands. *AgriEngineering*, *5*(2), 1051-1067. https://doi.org/10.3390/agriengineering5020066

Yang, C., Everitt, J., & Murden, D. (2011). Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture - COMPUT ELECTRON AGRIC*, *75*, 347-354. https://doi.org/10.1016/j.compag.2010.12.012

Zhou, J., Zan, M., Zhai, L., Yang, S., Xue, C., Li, R., & Wang, X. (2025). Remote sensing estimation of aboveground biomass of different forest types in Xinjiang based on machine learning. *Sci Rep*, *15*(1), 6187. https://doi.org/10.1038/s41598-025-90906-3

APPENDIX

Supplementary Materials

APPENDIX A

CHAPTER 3: MILK YIELD PREDICTION AND GRAIN-BASED OPTIMISATION CODE

This appendix presents key code excerpts from the model comparison and concentrate optimisation analysis conducted in Chapter 3. The complete implementation is provided in two files: (1) `modelling.ipynb` containing comparative evaluation of Random Forest, Neural Networks, LSTM, and Gaussian Process models for milk yield prediction; and (2) `scipy_optimisation.py` implementing Scipy's differential evolution algorithm for concentrate allocation optimisation. Both files are available in the project repository at: https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-3-data-driven-optimisation

File 1: Model Comparison Notebook

- Random Forest with grid search cross-validation
- Neural networks with Adam and SGD optimisers
- LSTM for time series prediction
- Gaussian Process regression
- Comparative performance evaluation

File 2: Scipy Differential Evolution Script

- Differential evolution optimisation algorithm
- Linear constraints for daily budget
- Dynamic bounds based on previous day
- Day-by-day optimisation over 91 days

A.1 Random Forest Model Training

The Random Forest model was trained with grid search cross-validation to identify optimal hyperparameters:

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV

# Define parameter grid
param_grid = {
    'n_estimators': [100, 500, 800],
```

```
    'max_depth': [10, 20, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

# Initialise model
rf = RandomForestRegressor(random_state=42)

# Grid search with cross-validation
grid_search = GridSearchCV(
    rf,
    param_grid,
    cv=5,
    scoring='neg_mean_squared_error',
    n_jobs=-1
)

# Train
grid_search.fit(X_train, y_train)

# Best model
best_rf = grid_search.best_estimator_
```

Source: model_comparison.ipynb, Random Forest section

A.2 Differential Evolution Optimisation

Scipy's differential evolution algorithm optimises concentrate allocation with linear constraints:

```
from scipy.optimize import differential_evolution, LinearConstraint
import numpy as np

# Define bounds for each cow
if i == 0:
    # First day: initial bounds
    bounds = [(5, 9)] * num_cows
else:
    # Subsequent days: dynamic bounds
    lower_bounds = np.maximum(optimal_conc_values_day - 1, 5)
    upper_bounds = np.minimum(optimal_conc_values_day + 1, 9)
    bounds = list(zip(lower_bounds, upper_bounds))

# Linear constraint: total concentrate ≤ budget
linear_constraint = LinearConstraint(
```

```
    np.ones((1, num_cows)),
    lb=[total_actual_conc * 0.95],
    ub=[total_actual_conc]
)

# Run optimisation
result = differential_evolution(
    objective_function,
    bounds,
    args=(current_day_df,),
    constraints=[linear_constraint]
)

# Extract optimal solution
optimal_conc_values_day = result.x
```

Source: scipy_optimisation.py, lines 122-152

A.3 Objective Function

The objective function predicts total milk yield for a given concentrate allocation:

```
def objective_function(x, data):
    """
    Calculate total predicted milk yield.

    Args:
        x: Concentrate allocation per cow
        data: Current day DataFrame

    Returns:
        float: Negative total milk yield (for minimisation)
    """
    # Update concentrate values
    data_copy = data.copy()
    data_copy['lact'] = data_copy['lact'].astype('object')
    data_copy['conc'] = x

    # Prepare features
    X = data_copy.drop(['id', 'my', 'smp_date'], axis=1)

    # Predict milk yield
    predicted_my = loaded_model.predict(X)
```

```
        # Append to tracking lists
        predictions.append(predicted_my)
        objective_values.append(-np.sum(predicted_my))

        # Return negative sum (for maximisation)
        return -np.sum(predicted_my)
```

Source: scipy_optimisation.py, lines 97-120

Note: *The complete implementation includes:*

- *Comprehensive model comparison across four algorithms*
- *Grid search hyperparameter optimisation*
- *Differential evolution with linear constraints*
- *Dynamic bounds for smooth day-to-day transitions*
- *Results export for 91 days of optimisation*

Repository access:

Model comparison:

https://github.com/Stansfash/thesis-python_codes/blob/main/chapter-3-data-driven-optimisation/Chapter_3_modelling.ipynb

Optimisation:

https://github.com/Stansfash/thesis-python_codes/blob/main/chapter-3-data-driven-optimisation/Chapter_3_Optimisation_code.py

APPENDIX B

CHAPTER 4: KEY CODE IMPLEMENTATIONS

This appendix presents key code excerpts from the machine learning pipeline and optimisation framework developed in Chapter 4 for MY prediction and optimisation of supplementary cow feed. Complete implementations of all scripts, including data loading, error handling, and auxiliary functions, are available in the project repository at: https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-4-machine-learning

B.1 Machine Learning Pipeline Construction

The following function creates the preprocessing and modeling pipeline used throughout Chapter 4. It automatically identifies numerical and categorical features, applies appropriate transformations, and combines them with the machine learning model.

```
def create_pipeline(model, X_train):
    """Creates ML pipeline with preprocessing steps."""
    numeric_features = X_train.select_dtypes(
        include=['int64', 'float64']).columns
    categorical_features = X_train.select_dtypes(
        include=['object']).columns

    numeric_transformer = StandardScaler()
    categorical_transformer = OneHotEncoder(
        handle_unknown='ignore')

    preprocessor = ColumnTransformer(
        transformers=[
            ('num', numeric_transformer, numeric_features),
            ('cat', categorical_transformer, categorical_features)
        ]
    )
    pipeline = Pipeline(steps=[
        ('preprocessor', preprocessor),
        ('model', model)
    ])
    return pipeline
```

Source: chapter-4-machine-learning/train_model.py, lines 100-125

B.2 Model Evaluation Metrics

The evaluation function calculates comprehensive performance metrics for model assessment:

```
def evaluate(true, predicted):
    """Calculate evaluation metrics."""
    mae = round(mean_absolute_error(true, predicted), 4)
    mape = round(mean_absolute_percentage_error(
        true, predicted), 4)
    mse = round(mean_squared_error(true, predicted), 4)
    msle = round(mean_squared_log_error(true, predicted), 4)
    rmse = round(np.sqrt(mse), 4)
    rmsle = round(np.sqrt(msle), 4)
    r2 = round(r2_score(true, predicted), 4)

    return mae, mape, mse, msle, rmse, rmsle, r2
```

Source: chapter-4-machine-learning/train_model.py, lines 128-147

B.3 Optimisation Problem Definition

The multi-objective optimisation problem defines decision variables (concentrate allocation per cow), objectives (maximise milk yield, minimise deviation), and constraints:

```
class DairyOptimisationProblem(ElementwiseProblem):
    def __init__(self, current_day_df, total_actual_conc,
                 loaded_model, previous_day_df=None):
        self.current_day_df = current_day_df
        self.total_actual_conc = total_actual_conc
        self.loaded_model = loaded_model
        self.previous_day_df = previous_day_df

        num_cows = len(current_day_df)
        bounds = [get_dynamic_bounds(cow_id, current_day_df,
                  previous_day_df)
                  for cow_id in current_day_df['cow_id']]
        xl, xu = zip(*bounds)

        super().__init__(n_var=num_cows, n_obj=2, n_constr=1,
                         xl=np.array(xl), xu=np.array(xu))

    def _evaluate(self, x, out, *args, **kwargs):
        self.current_day_df['conc'] = x

        X = self.current_day_df[["breed", "dim", "sol",
            "lact", "vp", "min_temp", "max_temp",
            "rh_tmin", "rh_tmax", "conc"]]
```

```
        predicted_my = self.loaded_model.predict(X)

        total_my = -np.sum(predicted_my)
        total_conc = np.sum(x)
        deviation_conc = np.abs(total_conc -
            self.total_actual_conc)
        constraint = [total_conc -
            (self.total_actual_conc * 0.999)]

        out["F"] = [total_my, deviation_conc]
        out["G"] = constraint
```

Source: chapter-4-machine-learning/multi_algorithm_optimization_fixed.py, lines 91-116

B.4 Dynamic Concentrate Bounds

To prevent abrupt dietary changes, concentrate allocation is constrained based on previous day's allocation (±2 kg), with absolute limits of 6-11 kg per cow:

```
def get_dynamic_bounds(cow_id, current_day_df, previous_day_df):
    """Determine bounds for concentrate allocation."""
    if (previous_day_df is not None and
        cow_id in previous_day_df['cow_id'].values):

        previous_conc = previous_day_df[
            previous_day_df['cow_id'] == cow_id
        ]['conc'].values[0]

        lower_bound = max(6, previous_conc - 2)
        upper_bound = min(11, previous_conc + 2)
    else:
        lower_bound = 6
        upper_bound = 11

    return lower_bound, upper_bound
```

Source: chapter-4-machine-learning/multi_algorithm_optimization_fixed.py, lines 64-84

B.5 Algorithm Initialisation

All four evolutionary algorithms (NSGA-II, SPEA2, SMS-EMOA, RVEA) are initialised with identical parameters to ensure fair comparison:

```
def get_algorithm(algorithm_name):
    """Initialize algorithm with identical parameters."""
    common_params = {
```

```
        'pop_size': 200,
        'sampling': FloatRandomSampling(),
        'crossover': SimulatedBinaryCrossover(
            eta=15, prob=0.95),
        'mutation': PolynomialMutation(eta=20),
        'eliminate_duplicates': True
    }

    if algorithm_name == 'NSGA2':
        return NSGA2(**common_params)
    elif algorithm_name == 'SPEA2':
        return SPEA2(**common_params)
    elif algorithm_name == 'SMSEMOA':
        return SMSEMOA(**common_params)
    elif algorithm_name == 'RVEA':
        ref_dirs = get_reference_directions(
            "das-dennis", n_dim=2, n_points=200)
        return RVEA(ref_dirs=ref_dirs, **common_params)
```

Source: chapter-4-machine-learning/multi_algorithm_optimization_fixed.py, lines 123-153

Repository files:

- *train_model.py (348 lines) - Complete model training script*
- *multi_algorithm_optimisation_fixed.py (404 lines) - Full optimisation framework*
- *statistical_analysis.py - Statistical comparison of algorithms*

APPENDIX C

CHAPTER 5: SENTINEL-2 PB PREDICTION IMPLEMENTATION CODE

This appendix presents key code excerpts from the Sentinel-2 based pasture biomass prediction analysis conducted in Chapter 5. The complete implementation is provided in two Jupyter notebooks: (1) `data_cleaning.ipynb` containing data preprocessing, merging of multiple data sources (plate meter, Sentinel-2, weather, grazing records), and creation of modelling datasets with various interpolation strategies; and (2) `modelling_visualisations.ipynb` containing model training, evaluation across multiple scenarios, and results visualisation. Both notebooks are available in the project repository at: https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-5-sentinel-biomass-prediction

Notebook 1: Data Cleaning and Preprocessing

- Merging plate meter, Sentinel-2, weather, and grazing data
- Temporal alignment and common date extraction
- Multiple interpolation strategies (linear, yearly, 2-year)
- Validation set creation (temporal and spatial splits)

Notebook 2: Modelling and Visualisations

- Feature selection impact analysis
- Interpolation method comparison
- Temporal transferability assessment
- Progressive training evaluation
- Farm-level and regional performance analysis

C.1 Data Merging and Temporal Alignment

The following approach demonstrates merging multiple data sources and extracting common temporal coverage:

```
# Extract common dates across all data sources
def extract_common_dates(plate_df, sentinel_df, weather_df):
    """
    Find dates present in all data sources.

    Returns:
```

```
        DataFrame: Merged data with common dates
    """
    # Get unique dates from each source
    plate_dates = set(plate_df['date'].unique())
    sentinel_dates = set(sentinel_df['date'].unique())
    weather_dates = set(weather_df['date'].unique())

    # Find intersection
    common_dates = plate_dates & sentinel_dates & weather_dates

    # Filter to common dates
    plate_common = plate_df[plate_df['date'].isin(common_dates)]
    sentinel_common = sentinel_df[sentinel_df['date'].isin(common_dates)]
    weather_common = weather_df[weather_df['date'].isin(common_dates)]

    # Merge on date and paddock
    merged = plate_common.merge(sentinel_common, on=['date', 'paddock'])
    merged = merged.merge(weather_common, on='date')

    return merged
```

Source: data_cleaning.ipynb, Common dates section

C.2 Interpolation Strategy

Multiple interpolation strategies were tested to handle missing Sentinel-2 observations due to cloud cover. Example linear interpolation implementation:

```
# Linear interpolation for NDVI time series
def interpolate_sentinel_data(df, method='linear'):
    """
    Interpolate missing Sentinel-2 values.

    Args:
        method: 'linear', 'yearly', or 'yearly2'
    """
    df_sorted = df.sort_values(['paddock', 'date'])

    sentinel_cols = ['NDVI', 'EVI', 'NDWI', 'SAVI']

    if method == 'linear':
        # Group by paddock and interpolate
        interpolated = df_sorted.groupby('paddock')[sentinel_cols].apply(
            lambda x: x.interpolate(method='linear', limit=30)
```

```
        )

    elif method == 'yearly':
        # Interpolate within each year only
        interpolated = df_sorted.groupby(['paddock', 'year'])[sentinel_cols].apply(
            lambda x: x.interpolate(method='linear')
        )

    return interpolated
```

Source: data_cleaning.ipynb, Interpolation section

C.3 Model Training and Evaluation

XGBoost models were trained with various feature combinations. Example training structure:

```
import xgboost as xgb
from sklearn.metrics import mean_squared_error, r2_score

# Prepare features
feature_sets = {
    'NDVI_only': ['NDVI'],
    'NDVI_weather': ['NDVI', 'temp_min', 'temp_max', 'rainfall'],
    'All_features': ['NDVI', 'EVI', 'SAVI', 'temp_min', 'temp_max', 'rainfall']
}

results = (NRC, 2001)

for name, features in feature_sets.items():
    # Train model
    model = xgb.XGBRegressor(
        n_estimators=100,
        max_depth=6,
        learning_rate=0.1
    )
    model.fit(X_train[features], y_train)

    # Evaluate
    y_pred = model.predict(X_test[features])

    results[name] = {
        'RMSE': np.sqrt(mean_squared_error(y_test, y_pred)),
        'R2': r2_score(y_test, y_pred)
    }
```

Source: modelling_visualisations.ipynb, Part 1

Note: *The complete implementation includes:*

- *Comprehensive data cleaning and merging pipelines*
- *Multiple interpolation strategies with comparative analysis*
- *Temporal and spatial validation approaches*
- *Feature selection and progressive training experiments*
- *Farm-level and regional performance visualisations*

Repository access:

Data cleaning:

`https://github.com/Stansfash/thesis-python_codes/blob/main/chapter-5-sentinel-biomass-prediction/Chapter_5_data_cleaning.ipynb`

Modelling:

`https://github.com/Stansfash/thesis-python_codes/blob/main/chapter-5-sentinel-biomass-prediction/Chapter_5_modelling_visualisations.ipynb`

APPENDIX D

CHAPTER 6: GRAZING EVENT DETECTION IMPLEMENTATION CODE

This appendix presents key code excerpts from the grazing event detection analysis conducted in Chapter 6. The complete implementation is provided as a Jupyter notebook containing data preprocessing, multiple detection algorithms (Blocks 1-12), model evaluation, and visualisation code. The full notebook (`Chapter_6_Grazing_Detection_Notebook.ipynb`), including all analysis blocks and figure generation code, is available in the project repository at: https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-6-grazing-detection

Notebook Structure:

- Library imports and data loading functions
- Detection algorithm blocks (Blocks 1-12)
- Statistical analysis and model evaluation
- Results visualisation (Figures 1-7)

D.1 Data Loading and Preprocessing

The following function demonstrates the data loading approach used throughout the analysis. It loads data from multiple folders and file patterns:

```
def load_files_from_folders(folders, patterns):
    """
    Loads files from specified folders matching patterns.

    Args:
        folders (list): List of folder paths
        patterns (list): File patterns to match

    Returns:
        dict: Dictionary of DataFrames by filename
    """
    dataframes = {}

    for folder in folders:
        for pattern in patterns:
            file_path = os.path.join(folder, pattern)
```

```
            matching_files = glob.glob(file_path)

            for file in matching_files:
                filename = os.path.basename(file)
                df = pd.read_csv(file)
                dataframes[filename] = df

    return dataframes
```

Source: Jupyter notebook, Cell 15

D.2 Grazing Detection Algorithm Structure

The grazing detection algorithms (Blocks 1-12) follow a consistent structure for model training, validation, and evaluation. Each block implements different combinations of features, temporal windows, and validation strategies. The general pattern for supervised learning blocks is:

```
# Example structure from detection blocks

# 1. Feature engineering and data preparation
X = df[feature_columns]
y = df['target_variable']

# 2. Train-test split (temporal or random)
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# 3. Model training
model = RandomForestClassifier(
    n_estimators=100,
    max_depth=10,
    random_state=42
)
model.fit(X_train, y_train)

# 4. Prediction and evaluation
y_pred = model.predict(X_test)

# 5. Performance metrics
from sklearn.metrics import (
    accuracy_score, precision_score,
    recall_score, f1_score
)
```

```
metrics = {
    'accuracy': accuracy_score(y_test, y_pred),
    'precision': precision_score(y_test, y_pred),
    'recall': recall_score(y_test, y_pred),
    'f1': f1_score(y_test, y_pred)
}
```

Source: Generalised pattern from Blocks 1-12

D.3 Results Visualisation

The notebook includes comprehensive visualisation code for presenting results. The figures use Plotly for interactive visualisations. Example structure for model comparison plots:

```
import plotly.graph_objects as go

# Prepare performance data
models = ['Model_1', 'Model_2', 'Model_3']
f1_scores = [0.89, 0.87, 0.85]

# Create bar chart
fig = go.Figure(data=[
    go.Bar(
        x=models,
        y=f1_scores,
        marker_color='rgb(55, 83, 109)'
    )
])

# Update layout
fig.update_layout(
    title='Model Performance Comparison',
    xaxis_title='Model',
    yaxis_title='F1 Score',
    showlegend=False
)

fig.show()
```

Source: Simplified from Figure generation cells

Note: *The complete Jupyter notebook contains:*

- *12 detection algorithm blocks with full implementations*

- *Comprehensive model evaluation across multiple metrics*
- *Statistical analysis and validation code*
- *Complete figure generation code for all 7 figures*

Repository access:

https://github.com/Stansfash/thesis-python_codes/blob/main/chapter-6-grazing-detection/Chapter_6_Grazing_Detection_Notebook.ipynb

APPENDIX E

CHAPTER 7: PB PREDICTION WITH IMAGES IMPLEMENTATION CODE

This appendix presents key code excerpts from the pasture biomass prediction analysis conducted in Chapter 7. The complete implementation is provided as a Jupyter notebook containing image preprocessing, feature extraction, model training (baseline and deep learning experiments), stacking ensemble methods, and results analysis. The full notebook (`Chapter_7_Biomass_Prediction_from_images_notebook.ipynb`), including all experiments (1-21), feature importance analysis, seasonal performance evaluation, and visualisation code, is available in the project repository at: https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-7-biomass-prediction-from-images

Notebook Structure:

- Part 1A: Exploratory data analysis and data loading
- Part 1B: Feature extraction and baseline experiments (Experiments 1-7)
- Part 2: Transfer learning experiments (Experiments 8-14)
- Part 3: Stacking and ensemble analysis (Experiments 15-21)
- Part 4: Results export and performance metrics
- Additional: Feature importance, seasonal analysis, biomass distribution

E.1 Image Feature Extraction

The following function extracts colour and texture features from pasture images for biomass prediction:

```
def extract_features(image_path):
    """
    Extract colour and texture features from image.

    Returns:
        dict: Feature dictionary with RGB, HSV stats
    """
    img = cv2.imread(image_path)
    img_rgb = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    img_hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)

    features = {}

    # RGB channel statistics
```

```
    for i, channel in enumerate(['R', 'G', 'B']):
        features[f'{channel}_mean'] = img_rgb[:,:,i].mean()
        features[f'{channel}_std'] = img_rgb[:,:,i].std()
        features[f'{channel}_median'] = np.median(img_rgb[:,:,i])

    # HSV channel statistics
    for i, channel in enumerate(['H', 'S', 'V']):
        features[f'{channel}_mean'] = img_hsv[:,:,i].mean()
        features[f'{channel}_std'] = img_hsv[:,:,i].std()

    return features
```

Source: Part 1B, Feature extraction section

E.2 Model Training and Evaluation

The baseline experiments (Experiments 1-7) compare multiple regression algorithms. Example training and evaluation structure:

```
# Prepare data
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
# Standardise features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Train Random Forest model
model = RandomForestRegressor(
    n_estimators=100,
    max_depth=20,
    random_state=42
)
model.fit(X_train_scaled, y_train)

# Evaluate performance
y_pred = model.predict(X_test_scaled)
metrics = {
    'RMSE': np.sqrt(mean_squared_error(y_test, y_pred)),
    'MAE': mean_absolute_error(y_test, y_pred),
    'R2': r2_score(y_test, y_pred)
}
```

Source: Generalised from Experiments 1-7

E.3 Stacking Ensemble Method

Experiments 15-21 implement stacking ensemble methods combining multiple base learners. Example stacking structure:

```
# Define base models
base_models = {
    'RF': RandomForestRegressor(n_estimators=100),
    'GBM': GradientBoostingRegressor(n_estimators=100),
    'Ridge': Ridge(alpha=1.0)
}
# Train base models and generate meta-features
meta_train = np.zeros((X_train.shape[0], len(base_models)))
meta_test = np.zeros((X_test.shape[0], len(base_models)))

for i, (name, model) in enumerate(base_models.items()):
    model.fit(X_train_scaled, y_train)
    meta_train[:, i] = model.predict(X_train_scaled)
    meta_test[:, i] = model.predict(X_test_scaled)

# Train meta-learner
meta_model = Ridge(alpha=1.0)
meta_model.fit(meta_train, y_train)

# Final predictions
y_pred_stack = meta_model.predict(meta_test)
```

Source: Simplified from Part 3, Experiments 15-21

Note: *The complete Jupyter notebook contains:*

- *21 complete experiments with full implementations*
- *Transfer learning experiments using pre-trained CNNs*
- *Feature importance analysis and visualisation*
- *Seasonal and regional performance evaluation*
- *Complete results export and statistical analysis*

Repository access:

https://github.com/Stansfash/thesis-python_codes/tree/main/chapter-7-biomass-prediction-from-images