THE UNIVERSITY OF
SYDNEY

PHD THESIS

DISCIPLINE OF BUSINESS ANALYTICS

GRAPH-BASED MACHINE LEARNING FOR REAL-WORLD APPLICATIONS

Author: **Kuan Yan**

Supervisor: Prof. Junbin Gao

A/Prof. Dmytro Matsypura

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

The University of Sydney

BUSINESS SCHOOL

2026

Contents

1 Introduction **9**

1.1 Background and Motivation 9

1.2 Graph-Based Machine Learning in Real-World Systems 11

1.3 Challenges in Applying Graph Learning to Practice 12

1.4 Graph-Based Learning for Finance and Medical AI 13

1.5 Thesis Contributions 16

1.6 Thesis Outline 19

2 Background **21**

2.1 Graph Representation in Machine Learning 21

2.1.1 Basic Concepts of Graphs 22

2.1.2 Graph Representation and Structural Information 23

2.1.3 Graph Learning Tasks and Applications 24

2.2 Graph Neural Networks 25

2.2.1 Message Passing Framework 26

2.2.2 Graph Convolutional Networks 27

2.2.3 Graph Representation Learning with GNNs 28

2.2.4 Applications of Graph Neural Networks 30

2.3 Fairness in Machine Learning and Graph Learning 31

2.3.1 Fairness in Machine Learning 31

2.3.2 Types of Fairness: Group Fairness and Individual Fairness . . 32

2.3.3 Fairness Challenges in Graph Neural Networks 33

2.4 Machine Learning in Biomedical Data Analysis 34
2.4.1 Machine Learning in Genomic and Transcriptomic Data 35
2.4.2 Challenges in Biomedical Data Analysis 36
2.4.3 Graph Learning for Biomedical Data Analysis 37

3 Heterogeneous Graph Learning for Credit Card Fraud Detection 39
3.1 Introduction 40
3.2 Related Work 42
3.2.1 Classic Machine Learning Approach 42
3.2.2 Deep Learning Approach 43
3.2.3 Graph Neural Network Approach 45
3.3 Methodology 46
3.3.1 CHET-graph 46
3.3.2 RGCN 50
3.3.3 Feature Importance-based Edge Weights 51
3.3.4 FIW-GNN 53
3.4 Experiments 54
3.4.1 Datasets Description 54
3.4.2 Compared Methods 56
3.4.3 Evaluation Metrics 57
3.4.4 Implementations and Parameter Settings 59
3.5 Experimental Results and Discussion 59
3.5.1 Performance Comparison 59
3.5.2 Ablation Study 63
3.6 Conclusions and Future Work 64

4 Balancing Group and Individual Fairness in Graph Neural Networks 67
4.1 Introduction 68
4.2 Related Work 70

4.2.1 Graph Neural Networks 71
4.2.2 Group Fairness in GNNs 72
4.2.3 Individual Fairness in GNNs 73
4.3 Preliminaries 75
4.3.1 Notations 75
4.3.2 Preliminaries of Graph Neural Networks 75
4.3.3 Balanced Fairness 76
4.3.4 Measuring Balanced Fairness 77
4.3.4.1 Measuring Group Fairness 77
4.3.4.2 Measuring Robust Individual Fairness 78
4.3.4.3 Balanced Fairness Score 80
4.3.5 Problem Definition 80
4.4 Proposed Framework 81
4.4.1 Framework Overview 81
4.4.2 Workflow of the Proposed Framework 82
4.4.2.1 Ensuring the Utility of the GNN Model 82
4.4.2.2 Enforcing Group Fairness 83
4.4.2.3 Enforcing Individual Fairness 83
4.4.3 Objective Function Formulation 84
4.5 Experiments 87
4.5.1 Datasets 88
4.5.2 Compared Methods 88
4.5.3 Evaluation Metrics 89
4.5.4 Effectiveness of BIG 90
4.5.5 Ablation study 92
4.6 Conclusions and Future Work 94

5 Graph-Based Machine Learning for Disease Severity Prediction and Progression Dynamics 96

5.1 Introduction . . . 97
5.2 Related Work . . . 100
5.2.1 Machine Learning in Biomedical Research . . . 100
5.2.2 RNA Sequencing Data in Machine Learning . . . 102
5.2.3 Molecular Studies on the Pathogenesis of Subretinal Fibrosis . 103
5.3 Methodology . . . 104
5.3.1 Framework Overview . . . 104
5.3.2 Animals . . . 107
5.3.3 Data Collection . . . 107
5.3.4 Feature Engineering . . . 109
5.3.4.1 Dimensionality Reduction . . . 109
5.3.4.2 Feature Expansion . . . 110
5.4 Experiments . . . 111
5.4.1 Dataset Description . . . 111
5.4.2 Prediction and Results . . . 111
5.4.2.1 Biological Correlation . . . 112
5.4.2.2 Influence Measurement . . . 113
5.4.3 Biological Impact Discussion . . . 115
5.5 Dynamic Modelling Extension via Graph Pseudotime Analysis and Neural SDEs . . . 118
5.5.1 Motivation . . . 118
5.5.2 Pathway Graph Construction . . . 118
5.5.3 Graph Pseudotime Analysis . . . 119
5.5.4 Neural SDE for Pathway Dynamics . . . 119
5.5.5 Pathway Stability . . . 120
5.5.6 Disease Bifurcation Points . . . 121
5.5.7 Discussion . . . 121
5.6 Conclusions . . . 122

6 Conclusions **123**

6.1 Summary of the Thesis 123

6.2 Key Contributions 125

6.3 Limitations and Future Work 126

List of Figures

3.1 An example of the CHET-graph. 49
3.2 Computation process for a node in CHET-graph using RGCN model. 51
3.3 The framework of proposed FIW-GNN method. 54
3.4 Histogram of the model evaluation metrics on the Vesta dataset. . . . 62
3.5 Histogram of the model evaluation metrics on the Sparkov dataset. . 63

4.1 The overall framework of BIG. 82
4.2 The pipeline of the BIG's optimization. 87
4.3 The effects of regularization on the performance of BIG. 92
4.4 Hyperparameter sensitivity analysis of BIG on the Census income dataset. 93

5.1 The overall framework of our study. 106
5.2 Key steps of the quantification method. 108
5.3 The logic of our iterative experiments. 112

List of Tables

3.1 Statistics of Datasets. 55

3.2 Performance comparison on the Vesta dataset for fraud detection. . . 60

3.3 Performance comparison on the Sparkov dataset for fraud detection. . 60

3.4 Evaluation metrics of different edge weight ranges in the ablation study. 65

4.1 Statistics of datasets. 89

4.2 Experimental results on three graph datasets. All entries are averages and standard deviations across five independent runs. Arrows $(\uparrow, \downarrow)$ indicate the direction of better performance. All individual fairness numbers are reported in tenths. IC stands for inconsistency. Best performances are shown in bold. 90

5.1 Top 7 most important genes selected by our ML models in biological correlation experiments with the greatest association to the disease. . 113

5.2 Top 10 genes selected by our ML models for their significant impact on disease improvement when gene expression values were reduced by 50%. 114

5.3 Top 10 genes selected by our ML models for their significant impact on disease exacerbation when gene expression values were increased to 200%. 115

CERTIFICATE OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Kuan Yan

_____

Abstract

Graph-based machine learning has emerged as a powerful paradigm for modeling complex real-world data, where entities and their interactions play a critical role. In many high-stakes application domains, such as finance and medical research, data are often heterogeneous, noisy, incomplete, and structurally complex, posing significant challenges to traditional machine learning approaches. This thesis investigates how graph-based machine learning methods can be designed and applied to address these challenges in real-world financial and medical AI applications, with an emphasis on practical modeling, robustness, interpretability, and domain relevance.

The thesis first focuses on financial applications, where graph-based representations are constructed to capture complex transactional relationships in credit card fraud detection. A heterogeneous graph learning framework is proposed to effectively utilize raw transactional data while remaining robust to missing values and noise. By incorporating feature importance–based edge weighting within a graph neural network architecture, the proposed approach achieves improved detection performance and enhanced interpretability compared to existing methods. Building on graph-based modeling in financial systems, the thesis further explores fairness considerations in graph neural networks. A practical framework is introduced to balance group fairness and individual fairness within graph-based learning, demonstrating that these fairness objectives need not be inherently conflicting in real-world applications.

The second part of the thesis shifts focus to medical AI, where graph-based and machine learning methods are applied to analyze high-dimensional

biological data related to retinal disease. A machine learning framework is developed to predict subretinal lesion severity and identify key genes associated with disease progression in a mouse model of age-related macular degeneration, addressing challenges such as limited sample sizes and high dimensionality in transcriptomic data. Extending beyond static prediction, the thesis further investigates disease progression dynamics through graph-level modeling of biological pathways. A graph pseudotime analysis framework is proposed to infer population-level disease trajectories, and neural stochastic differential equations are employed to characterize pathway stability and identify critical disease transition points.

Across the financial and medical domains, this thesis demonstrates that graph-based machine learning provides a flexible and effective framework for modeling complex real-world data, enabling robust prediction, fair decision-making, and improved interpretability. The findings provide evidence of the potential value of graph-based approaches for high-risk applications in finance and medical AI, while further validation and investigation are required before large-scale real-world deployment.

Acknowledgements

This thesis marks the culmination of my PhD journey and reflects the collective support, guidance, and encouragement I have received along the way. I am deeply grateful to all those who have accompanied me throughout this journey, both academically and personally, during its challenges and achievements.

First and foremost, I would like to express my deepest gratitude to my principal supervisor, Professor Junbin Gao, for his invaluable guidance, patience, and unwavering support throughout my doctoral studies. His profound academic insight, rigorous standards, and open-door mentorship have played a pivotal role in shaping my research direction and intellectual development. His encouragement and trust have continually motivated me to pursue meaningful and impactful research, and his guidance has been instrumental in helping me navigate the challenges of the PhD journey.

I would also like to sincerely thank my associate supervisor, A/Prof. Dmytro Matsypura, for his constructive feedback, thoughtful advice, and consistent support. His perspectives and suggestions have been highly valuable in refining my research and strengthening the quality of this thesis.

I am grateful to Professor Mark Gillies and Dr. Ling Zhu for their generous collaboration and academic support. Their expertise, insights, and willingness to share knowledge have greatly enriched my research

experience and contributed meaningfully to the interdisciplinary aspects of my work.

I gratefully acknowledge the financial support provided by the Enhanced Business School Research Scholarship (EBSRS) at The University of Sydney Business School. This scholarship has enabled me to undertake my doctoral studies with stability and focus, and to fully dedicate myself to the research presented in this thesis.

I would also like to thank my colleagues and friends for their support, encouragement, and companionship throughout my PhD journey. Their presence and shared experiences have made this journey both rewarding and memorable.

Finally, I am profoundly grateful to my family for their unconditional love, understanding, and unwavering support. Their constant encouragement has been a source of strength throughout my doctoral studies and has given me the confidence to persevere through challenges. Without them, this journey would not have been possible.

Authorship Attribution Statement

This thesis was conducted at The University of Sydney, under the supervision of Prof. Junbin Gao and A/Prof. Dmytro Matsypura, between 2022 and 2026. For the works included in this thesis, I made the main contributions by conceiving the idea, designing the research, conducting the experiments, and drafting the paper.

This thesis is based on the following publications:

- K. Yan, J. Gao and D. Matsypura, (2023). FIW-GNN: A Heterogeneous Graph-Based Learning Model for Credit Card Fraud Detection. *Proceedings of the IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA 2023)*, pp. 1-10.
- K. Yan, D. Matsypura and J. Gao, (2024). BIG: A Practical Framework for Balancing the Conflict Between Group and Individual Fairness in Graph Neural Networks. *Proceedings of the IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2024)*, pp. 2400-2409.
- K. Yan, Y. Zeng, D. Shi, T. Zhang, D. Matsypura, M. C. Gillies, L. Zhu and J. Gao, (2025). Machine Learning-Based Prediction of Key Genes Correlated to the Subretinal Lesion Severity in a Mouse Model of Age-Related Macular Degeneration. *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025)*, Volume 1, pages 627-637.

The author has also made substantial contributions to the following research works that are not included in this thesis, for which the author is the co-first author and has an equal contribution with the first author.

- D. Shi*, K. Yan*, L. Lin*, Y. Zeng, T. Zhang, D. Matsypura, M. C. Gillies, L. Zhu and J. Gao, (2025). Graph Pseudotime Analysis and Neural Stochastic Differential Equations for Analyzing Retinal Degeneration Dynamics and Beyond. *Proceedings of ICLR 2025 Workshop on Machine Learning for Genomics Explorations (MLGENX), International Conference on Learning Representations (ICLR 2025)*, arXiv preprint arXiv:2502.06126.

In addition to the statements above, in cases where Kuan (Michael) Yan is not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Kuan Yan 27/3/2026

As a supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Junbin Gao 27/3/2026

Use of Generative Artificial Intelligence

The author used generative artificial intelligence to improve the clarity and presentation of the writing of this thesis. Any edits by generative artificial intelligence were carefully reviewed. All ideas, research, analyses, and conclusions are entirely the author's own.

Kuan Yan

Chapter 1

Introduction

1.1 Background and Motivation

In recent years, the rapid development of machine learning and data-driven methods has fundamentally transformed how complex real-world problems are analyzed and addressed (Sharifani & Amini, 2023; Shehab et al., 2022). With the increasing availability of large-scale data and computational resources, learning-based models have become central to decision-making systems across a wide range of domains, including finance, healthcare, and scientific research.

Despite these advances, many real-world problems cannot be adequately represented as collections of independent data points. Instead, they are inherently structured, involving rich relationships and interactions among entities (Veličković, 2023). Examples include financial transaction networks that connect accounts and payments, biological systems characterized by interactions between genes and proteins, and clinical settings in which patients, diseases, and treatments are tightly interconnected. Ignoring such relational dependencies may lead to incomplete representations and, consequently, to suboptimal decision-making.

Graph-based data representations provide a principled way to explicitly model these relational structures (L. Wu et al., 2022; C. Gao, Zheng, et al., 2023). By representing entities as nodes and their interactions as edges, graphs provide a flexible

and expressive framework for capturing complex system structures that are difficult to model using standard feature-based representations. As a result, graph-based machine learning has emerged as an effective paradigm for learning from relational data and supporting downstream tasks such as prediction, classification, and decision support.

In application-driven domains such as finance and medical artificial intelligence, the importance of relational modeling is particularly pronounced. Financial systems are shaped by intricate transaction patterns, behavioral dependencies, and evolving risk structures, where inaccurate predictions can lead to substantial financial losses or systemic risks (Motie & Raahemi, 2024). Similarly, medical and biomedical data often involve multi-level relationships among biological entities, patients, and clinical outcomes (Paul et al., 2024; C. Gao, Yin, et al., 2023), in which erroneous assessments or decisions can result in serious clinical consequences. In these high-stakes settings, effective decision-making critically depends on the ability to capture underlying relational structures rather than rely solely on isolated features (Bing et al., 2023). These characteristics naturally motivate the adoption of graph-based learning approaches as a principled and effective means of extracting meaningful patterns from complex real-world data.

Motivated by these observations, this thesis focuses on graph-based machine learning methods for real-world applications, with a particular emphasis on financial systems and medical artificial intelligence. Rather than viewing graphs purely as a formal representation, this thesis investigates how graph-based learning methods can be practically designed and applied to address domain-specific challenges arising in realistic data environments.

1.2 Graph-Based Machine Learning in Real-World Systems

The increasing prevalence of relational data has led to the rapid development of graph-based machine learning methods, particularly graph neural networks (GNNs), as a powerful framework for modeling complex structured systems (Z. Wu et al., 2020). By integrating node features with topological information through iterative aggregation mechanisms, GNNs are able to learn representations that capture both local interactions and global structural patterns (Zhou et al., 2020).

In financial systems, graph-based learning has become especially valuable due to the inherently networked nature of transactions and interactions. Financial activities such as payment transfers, credit relationships, and account interactions naturally form transaction graphs, where suspicious behavior often manifests itself through relational patterns rather than isolated attributes. Graph-based models have been widely explored for fraud detection, risk assessment, anti-money laundering, and credit scoring, demonstrating the advantage of relational modeling in identifying subtle structural anomalies and behavioral dependencies (Innan et al., 2024; Cheng et al., 2025).

Similarly, in medical and biomedical domains, relational structures are equally fundamental. Instead of existing as isolated measurements, biological and clinical entities are interconnected through complex interaction mechanisms. For example, genes and proteins operate within interaction networks, and diseases are often associated with groups of molecular components rather than single factors (Y. Li et al., 2022). At the clinical level, patients can also be modeled through similarity networks that integrate diagnostic records, treatment histories, and demographic attributes. These interconnected representations enable graph-based learning methods to support tasks such as disease prediction, biomarker discovery, and treatment outcome modeling (Sun et al., 2022; Pfeifer et al., 2022).

Beyond finance and healthcare, graph-based machine learning has been increasingly applied in a broad range of other domains where relational dependencies play a central role. Examples include recommendation systems in e-commerce platforms, social network analysis, transportation systems, knowledge graphs, and communication networks (Yang et al., 2023; G. Zhang et al., 2022; Sharma et al., 2024). In these domains, meaningful performance gains are often driven by the ability to model interactions among entities, uncover structural patterns, and account for long-range dependencies across the network.

Collectively, these examples highlight that graph-based machine learning is not merely an extension of conventional deep learning models, but a necessary framework for analyzing interconnected systems (Z. Song et al., 2022). Across diverse application domains, relational modeling enhances the ability to extract meaningful patterns from complex data structures and supports more informed and reliable decision-making processes.

1.3 Challenges in Applying Graph Learning to Practice

Despite the growing success of graph-based methods, their deployment in real-world environments remains challenging. Practical data conditions often deviate from idealized benchmark settings, introducing complexities that may limit model generalization and reliability (Khemani et al., 2024; Munikoti et al., 2023).

First, real-world graph data are frequently heterogeneous and noisy (C.-T. Li et al., 2025). Financial networks may involve diverse entity types and evolving behavioral patterns, while medical graphs often integrate multi-modal biological and clinical data with varying scales and uncertainties (R. Li et al., 2021). Such heterogeneity complicates representation learning and requires models to balance structural expressiveness with stability.

Second, labeled data are often scarce or highly imbalanced in realistic settings. In fraud detection, anomalous transactions typically represent a small fraction of the overall population (Ali et al., 2022). In biomedical studies, labeled disease samples or experimentally validated interactions may be limited due to cost and ethical constraints (Moser et al., 2023). Learning robust representations under such imbalance poses significant methodological and practical challenges.

Third, real-world systems are dynamic. Financial networks evolve as new accounts and transactions are introduced, and medical knowledge continuously expands with emerging clinical evidence. Distribution shifts over time may degrade model performance if learned representations fail to adapt.

These challenges highlight a critical gap between theoretical advancements in graph neural networks and their practical deployment in complex application domains (Ju et al., 2025). While increasingly sophisticated architectures have been proposed to improve benchmark performance, real-world effectiveness often depends not only on model complexity but also on principled design choices that account for data characteristics, domain constraints, and decision-making requirements (Mohi ud din & Qureshi, 2023).

1.4 Graph-Based Learning for Finance and Medical AI

While graph-based machine learning has found applications across numerous domains, finance and medical artificial intelligence represent two particularly demanding and impactful settings (Pham et al., 2025). These domains are characterized not only by rich relational structures, but also by stringent requirements on reliability, robustness, interpretability, and real-world deployability (Y. Wang et al., 2025). As such, they provide a compelling context for examining how graph-based learning can be designed to operate effectively under practical constraints.

In financial systems, decision-making processes are tightly coupled with risk management and regulatory oversight. Models are often required to operate at large scale, process continuously evolving transaction streams, and detect rare yet consequential events (Zakaria et al., 2025). Beyond predictive accuracy, practical considerations such as computational efficiency, stability under distribution shifts, and transparency of model behavior are essential. A graph-based approach in this context must therefore balance expressive power with scalability, and structural sensitivity with robustness to noise and adversarial manipulation (Cui et al., 2025). Designing learning mechanisms that can effectively leverage relational signals while remaining stable in dynamic financial environments is a non-trivial challenge (Han et al., 2025).

A representative example is credit card fraud detection, where transactions can be modeled as interconnected entities linked through shared devices, merchants, geographic locations, or behavioral similarities. Fraudulent activities often manifest through coordinated or relational patterns rather than through obvious attribute anomalies (J. Wang et al., 2025). In such settings, graph-based learning enables the detection of subtle structural irregularities that may not be observable when analyzing transactions independently. At the same time, the extreme class imbalance and evolving fraud strategies require models that are both adaptive and robust, highlighting the importance of principled design choices beyond purely architectural complexity (Boyapati & Aygun, 2025).

Medical and biomedical applications introduce a different yet equally demanding set of constraints. Clinical decision-making directly affects patient outcomes, and errors may carry significant ethical and societal implications (Y. Gao et al., 2025). Data are frequently heterogeneous, sparse, and multi-modal, spanning molecular measurements, imaging data, electronic health records, and demographic information. Moreover, labeled data may be limited due to cost, privacy, and ethical considerations. In such environments, graph-based models must not only capture complex biological and clinical relationships, but also generalize reliably across cohorts and

settings (Gu et al., 2025). Emphasis is often placed on robustness, interpretability, and the ability to integrate prior domain knowledge into the learning process.

For example, in disease risk prediction or severity assessment, patients may be connected through similarity networks constructed from clinical features, genetic markers, or imaging-derived representations (Jin et al., 2025). Modeling such relationships allows graph-based approaches to incorporate contextual information from related cases, potentially improving predictive reliability. In molecular-level studies, identifying disease-associated gene modules or interaction substructures often benefits from analyzing the topology of biological networks rather than examining genes in isolation (Alharbi et al., 2025). These examples illustrate how relational modeling can contribute to more informed and clinically meaningful analyses.

Despite their differences, finance and medical AI share several underlying characteristics that make them particularly suitable for studying application-oriented graph learning. Both domains involve interconnected systems in which relationships are often more informative than isolated attributes. Both operate in high-stakes environments where decisions carry significant consequences. Both face challenges related to data imbalance, evolving distributions, and the need for principled model design beyond purely benchmark-driven optimization (Feng et al., 2025).

These shared characteristics motivate a focused investigation into how graph-based learning frameworks can be systematically tailored to meet domain-specific requirements in finance and medical artificial intelligence (Nandan et al., 2025). Rather than pursuing increasingly complex architectures in isolation, this thesis explores how objective formulation, data utilization strategies, and architectural considerations can be aligned with the structural and practical realities of these two domains. By grounding methodological development in realistic application settings, this work aims to advance graph-based learning toward more reliable and deployable solutions in high-impact environments.

1.5 Thesis Contributions

Building upon the application-oriented perspective established in the preceding sections, this thesis is guided by a single overarching research objective: To investigate how graph-based machine learning systems can be systematically designed to operate effectively under the practical constraints of high-stakes real-world applications.

Rather than focusing on a single application domain or a single graph learning methodology, this objective is addressed through three complementary research studies. These studies investigate application-oriented graph learning from different perspectives, including graph representation learning for noisy relational data, fairness-aware objective design for graph neural networks, and machine learning and graph-based modeling for biomedical intelligence. Although each study addresses a distinct research problem, they collectively contribute to a common objective of developing graph-based learning systems that are more robust, interpretable, trustworthy, and applicable to complex real-world environments.

The contributions of this thesis are therefore organized around these three complementary research studies.

The first contribution of this thesis lies in developing a heterogeneous graph-based learning framework for credit card fraud detection. Real-world transactional datasets are characterized by missing values, noisy attributes, and evolving behavioral patterns, which limit the effectiveness and stability of traditional machine learning and deep learning approaches.

To address these challenges, this work proposes the Credit-card HEtero Transactional graph (CHET-graph), a novel heterogeneous graph construction strategy tailored to credit card transactional data. By modeling transactions, accounts, and related entities within a unified relational structure, the proposed framework enables more comprehensive utilization of raw data without being severely affected by missing values. Furthermore, a feature-importance-based edge weighting mechanism is introduced to enhance interpretability and structural expressiveness. These compo-

nents are integrated into a relational graph convolutional architecture, resulting in the FIW-GNN model.

Through extensive empirical evaluations on benchmark transactional datasets, this study demonstrates that principled graph construction and feature-aware relational modeling can significantly improve fraud detection performance while maintaining robustness under realistic constraints. This contribution establishes a practical foundation for deploying graph-based learning in financial risk management systems.

The second contribution addresses fairness considerations in graph neural networks within high-stakes application environments. In domains such as finance and healthcare, predictive performance alone is insufficient; models must also avoid unfair treatment across demographic groups and similar individuals.

This study challenges the conventional view that group fairness and individual fairness are inherently conflicting objectives. Instead, it introduces the notion of balanced fairness and proposes the BIG (Balancing Individual and Group fairness) framework to jointly optimize both fairness perspectives. A novel evaluation metric, the Balanced Fairness Score, is designed to quantify the equilibrium between these objectives.

By demonstrating that group and individual fairness can be effectively balanced without sacrificing predictive performance, this research contributes a practical fairness-aware graph learning framework. It highlights that relational models deployed in sensitive domains must integrate fairness considerations into their design, thereby extending graph-based learning toward more socially responsible and deployable systems.

The third contribution of this thesis focuses on medical artificial intelligence, particularly the modeling of disease severity and progression dynamics in age-related macular degeneration (AMD). Unlike static classification problems, medical decision-making often requires understanding both the molecular determinants of disease severity and the temporal evolution of pathological processes.

First, this research develops a machine learning framework for predicting lesion severity and identifying key genes associated with subretinal fibrosis using high-dimensional RNA sequencing data from a mouse model. By integrating pathway-informed dimensionality reduction and gene-level feature expansion with Ridge and ElasticNet regression models, the study addresses challenges related to limited sample sizes and high dimensionality while maintaining biological interpretability. The proposed approach identifies potential therapeutic gene targets, bridging transcriptomic analysis and translational medicine.

Building upon this gene-level analysis, the research further advances toward modeling disease dynamics at the pathway level. A biologically informed graph-forming strategy is developed to construct pathway graphs from transcriptomic data, and Graph-level Pseudotime Analysis (GPA) is proposed to infer population-level disease trajectories. Neural stochastic differential equations (SDEs) are then introduced to characterize pathway stability and detect disease bifurcation points. By integrating graph-based structural modeling with dynamical system analysis, this study extends relational learning beyond static prediction toward mechanistic understanding of disease evolution.

Collectively, these three research studies demonstrate that graph-based machine learning can serve as a principled and versatile framework for addressing interconnected decision problems in finance and medical AI. Across fraud detection, fairness-aware modeling, and disease progression analysis, this thesis illustrates how relational modeling, objective design, and domain-informed methodological development can enhance robustness, interpretability, and practical deployability in high-impact environments.

By grounding technical advances in realistic application settings, this work contributes toward a more application-oriented paradigm of graph-based learning. The findings provide evidence of the potential value of graph-based approaches for high-stakes applications in finance and medical AI, while further validation and large-scale real-world deployment remain important directions for future research.

1.6 Thesis Outline

This thesis is organized into six chapters, with the main content of each chapter briefly described as follows.

Chapter 1: Introduction

This chapter introduces the research background and motivation, emphasizing the importance of relational modeling in high-stakes domains such as finance and medical artificial intelligence. It outlines the practical challenges of applying graph-based learning in real-world environments and summarizes the main contributions of the thesis.

Chapter 2: Background

This chapter reviews the theoretical foundations of graph-based machine learning, including fundamental concepts of graph representation learning, graph neural networks (GNNs), and related relational learning paradigms. It establishes the methodological preliminaries that underpin the subsequent research studies presented in this thesis.

Chapter 3: Heterogeneous Graph Learning for Credit Card Fraud Detection

This chapter presents a heterogeneous graph-based learning framework for credit card fraud detection. It introduces the construction of the Credit-card HEtero Transactional graph (CHET-graph) and a feature importance-based edge weighting mechanism integrated within a relational graph convolutional architecture. The proposed FIW-GNN model is evaluated on benchmark datasets, demonstrating its effectiveness and robustness under realistic transactional data constraints.

Chapter 4: Balancing Group and Individual Fairness in Graph Neural Networks

This chapter investigates fairness issues in graph neural networks within high-stakes

applications. It introduces the BIG framework, which jointly optimizes group fairness and individual fairness through the concept of balanced fairness and a novel evaluation metric. Empirical evaluations demonstrate that fairness objectives can be effectively balanced without sacrificing predictive performance.

Chapter 5: Graph-Based Machine Learning for Disease Severity Prediction and Progression Dynamics

This chapter focuses on medical artificial intelligence, presenting a unified framework that integrates gene-level prediction and pathway-level dynamic modeling. It first introduces a machine learning approach for predicting lesion severity and identifying key genes associated with subretinal fibrosis using RNA sequencing data. It then extends the analysis to pathway-level modeling through Graph-level Pseudotime Analysis (GPA) and neural stochastic differential equations, enabling the characterization of disease trajectories, pathway stability, and bifurcation dynamics.

Chapter 6: Conclusions

This chapter concludes the thesis by summarizing the main findings and contributions of the three research studies. It discusses the broader implications of application-oriented graph-based learning in finance and medical AI and outlines potential future research directions toward more robust, interpretable, and deployable relational learning systems.

Chapter 2

Background

2.1 Graph Representation in Machine Learning

Graphs have become an essential tool for representing complex systems composed of interacting entities. In many real-world scenarios, the relationships between entities are as important as the entities themselves. For example, financial transaction systems can be naturally represented as networks where accounts are connected through payment interactions, biological systems often involve networks of genes or proteins that regulate one another, and social systems consist of individuals linked through various forms of relationships. These relational structures cannot be fully captured using traditional tabular data representations because the dependencies among data points play a critical role in understanding the underlying system. Graph-based representations provide a natural and flexible framework for modeling such relational data.

In machine learning, graphs enable models to capture both the attributes of individual entities and the structural relationships among them. By representing data as interconnected nodes and edges, graph-based models can leverage structural information to improve predictive performance and uncover patterns that are difficult to detect using conventional methods. As a result, graph representations have been widely used in applications such as fraud detection, recommendation systems,

social network analysis, and biomedical research. To formally describe graphs used in machine learning tasks, the following sections introduce the basic concepts of graphs, their mathematical representations, and common learning tasks defined on graph-structured data.

2.1.1 Basic Concepts of Graphs

Mathematically, a graph can be defined as a tuple

$$G = (V, E) \tag{2.1}$$

where $V = \{v_1, v_2, ..., v_{|V|}\}$ denotes the set of nodes (or vertices), and E represents the set of edges that describe relationships between nodes. Each edge typically connects two nodes and can be written as a pair (v_i, v_j), indicating that node v_i is related to node v_j. The number of nodes in the graph is denoted by $|V|$, while $|E|$ represents the total number of edges.

Nodes represent entities within the system being modeled, while edges capture interactions or relationships between these entities. For example, in a financial transaction network, nodes may represent bank accounts and edges represent transactions between accounts. In a biological network, nodes may correspond to genes or proteins, and edges may represent regulatory interactions or functional relationships.

A key concept in graph analysis is the neighborhood of a node. Two nodes that are directly connected by an edge are referred to as one-hop neighbors. The set of neighbors of a node v is typically denoted as

$$N(v) = \{u \in V \mid (u, v) \in E\} \tag{2.2}$$

which contains all nodes that share a direct connection with v. Nodes that can be reached through two consecutive edges are referred to as two-hop neighbors, and this notion can be extended to higher-order neighborhoods.

Graphs can be categorized into different types depending on the characteris-

tics of their nodes and edges. For example, graphs may be directed or undirected depending on whether edges have orientations. Graphs can also be weighted or unweighted depending on whether edges carry numerical values representing the strength of interactions. In addition, graphs may be homogeneous when they contain a single type of node and edge, or heterogeneous when multiple types of entities or relationships exist.

2.1.2 Graph Representation and Structural Information

In machine learning applications, graphs are typically represented using mathematical structures that encode both attribute information and structural relationships. Two commonly used representations are the node feature matrix and the adjacency matrix.

The attribute information associated with nodes is usually represented by a feature matrix

$$X \in \mathbb{R}^{N \times F} \tag{2.3}$$

where $N = |V|$ denotes the number of nodes in the graph and F represents the number of features associated with each node. Each row of the matrix corresponds to the feature vector of a node, capturing properties such as demographic attributes, transaction statistics, or biological measurements depending on the application domain.

In addition to node attributes, the structure of the graph is typically encoded using an adjacency matrix

$$A \in \mathbb{R}^{N \times N} \tag{2.4}$$

where each entry A_{ij} describes the relationship between node v_i and node v_j. In an unweighted graph, $A_{ij} = 1$ indicates the presence of an edge between the two nodes, while $A_{ij} = 0$ indicates the absence of a connection. In weighted graphs,

A_{ij} may take real values representing the strength, frequency, or importance of the relationship.

The adjacency matrix provides a convenient mathematical representation of graph structure and allows many graph-based algorithms to be implemented efficiently using matrix operations. For undirected graphs, the adjacency matrix is symmetric such that $A_{ij} = A_{ji}$. In contrast, directed graphs generally produce asymmetric adjacency matrices.

Combining node attributes and structural information, a graph can be represented as

$$G = (X, A) \tag{2.5}$$

where X captures node-level features and A describes the connectivity structure among nodes. This representation forms the foundation of many graph-based machine learning models, enabling algorithms to jointly leverage node features and structural relationships when learning representations from graph data.

2.1.3 Graph Learning Tasks and Applications

Based on the graph representation described above, a variety of machine learning tasks can be formulated on graph-structured data. These tasks differ depending on the level at which predictions are made within the graph.

One of the most common tasks is node classification, where the objective is to predict the label of each node in the graph. In financial transaction networks, for instance, node classification can be used to identify potentially fraudulent accounts based on both their attributes and their connections to other accounts. Similarly, in citation networks, node classification can be applied to determine the research field of academic papers according to their citation relationships.

Another important task is link prediction, which focuses on estimating the likelihood that a connection exists between two nodes. This task is widely used in

recommendation systems and social network analysis. For example, link prediction can be used to recommend potential friends in a social network or to identify possible interactions between biological molecules in biomedical networks.

Beyond node-level and edge-level tasks, machine learning can also be applied at the graph level. Graph-level prediction, often referred to as graph classification or graph regression, aims to infer properties associated with an entire graph. This type of task frequently appears in applications such as molecular property prediction, where each graph represents a molecule and the objective is to estimate chemical or biological properties.

Graph-based machine learning has therefore been widely adopted in many domains. In financial systems, graph representations enable the modeling of complex transaction patterns that may reveal fraudulent activities. In recommendation systems, graphs capture relationships between users and items. In biomedical research, biological interaction networks can be analyzed to better understand disease mechanisms and identify potential biomarkers.

Although graph representations provide a powerful framework for modeling relational data, traditional machine learning models are not designed to directly process graph structures. This limitation has motivated the development of specialized models capable of learning from graph-structured data. Among these approaches, graph neural networks (GNNs) have emerged as one of the most effective techniques for learning representations from graphs. The following section introduces the fundamental principles of graph neural networks.

2.2 Graph Neural Networks

While graphs provide a natural way to represent complex relational data, traditional machine learning models are not well suited for directly processing graph-structured inputs. Conventional algorithms typically assume that data instances are independent and represented in fixed-length feature vectors. However, in many real-world

systems, entities are interconnected and their relationships play an essential role in determining their properties. Ignoring these relationships may lead to the loss of important structural information.

Graph neural networks (GNNs) have emerged as a powerful framework for learning from graph-structured data. The key idea behind GNNs is to learn node representations by aggregating information from neighboring nodes and their connections. Through iterative neighborhood aggregation, each node progressively incorporates information from a larger portion of the graph, allowing the model to capture both local interactions and global structural patterns.

Since their introduction, GNNs have achieved remarkable success in a wide range of applications, including social network analysis, recommender systems, financial risk detection, and biomedical data analysis. Compared with traditional approaches, GNNs are capable of jointly modeling node attributes and graph topology, enabling more expressive representations of relational data.

Most modern GNN architectures can be interpreted under a unified message passing framework, where nodes exchange information with their neighbors and update their representations through iterative aggregation and transformation operations. The following sections introduce the fundamental principles of message passing and discuss several representative GNN models.

2.2.1 Message Passing Framework

Most modern graph neural networks can be described within a unified message passing framework. The central idea of this framework is that each node updates its representation by aggregating information from its neighboring nodes. Through multiple rounds of information propagation, nodes are able to capture increasingly larger structural contexts in the graph.

Let $h_v^{(k)}$ denote the representation of node v at the k-th layer of a GNN. The message passing process typically consists of two main steps: neighborhood aggre-

gation and node update. During the aggregation step, a node collects information from its neighbors. During the update step, the node combines the aggregated information with its current representation to produce a new embedding.

Formally, the update process can be expressed as

$$h_v^{(k)} = \sigma\left(W^{(k)} \cdot \mathrm{AGG}\left(\{h_u^{(k-1)} : u \in N(v)\}\right)\right) \tag{2.6}$$

where $N(v)$ denotes the set of neighbors of node v, $W^{(k)}$ represents the learnable weight matrix at layer k, $\mathrm{AGG}(\cdot)$ denotes an aggregation function such as mean, sum, or max pooling, and $\sigma(\cdot)$ is a nonlinear activation function.

By stacking multiple message passing layers, each node is able to incorporate information from nodes that are multiple hops away in the graph. For example, after K layers of propagation, the representation of a node contains information from its K-hop neighborhood. This mechanism allows GNNs to effectively capture complex relational patterns in graph-structured data.

The message passing framework provides a general perspective that unifies many different GNN architectures. Various models differ mainly in the choice of aggregation functions, normalization strategies, and update mechanisms. Among these models, Graph Convolutional Networks (GCNs) represent one of the most influential and widely adopted approaches, which will be introduced in the next subsection.

2.2.2 Graph Convolutional Networks

Among the many graph neural network architectures proposed in recent years, Graph Convolutional Networks (GCNs) represent one of the most influential and widely adopted models. GCNs extend the idea of convolution from grid-structured data, such as images, to graph-structured data. Instead of applying convolution over spatially ordered pixels, graph convolution operates over irregular graph neighborhoods, enabling nodes to aggregate information from their connected neighbors.

The Graph Convolutional Network model proposed by Kipf and Welling (2017) provides a simple yet effective formulation for graph representation learning. In this model, node representations are updated through a normalized aggregation of neighboring features. Let $H^{(l)}$ denote the matrix of node representations at layer l, where each row corresponds to the embedding of a node. The propagation rule of a GCN layer can be written as

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{2.7}$$

where $\tilde{A} = A + I$ is the adjacency matrix with added self-loops, I is the identity matrix, $\tilde{D}$ is the corresponding degree matrix defined as $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, $W^{(l)}$ denotes the trainable weight matrix of layer l, and $\sigma(\cdot)$ represents a nonlinear activation function such as ReLU.

The normalization term $\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ plays an important role in stabilizing the aggregation process. Without normalization, nodes with many neighbors could dominate the aggregation results, potentially leading to numerical instability. The symmetric normalization ensures that information from neighboring nodes is appropriately scaled during the propagation process.

Through multiple layers of graph convolution, each node gradually integrates information from nodes that are multiple hops away in the graph. This process allows GCNs to capture both node-level attributes and structural dependencies within the graph. Due to their simplicity and strong empirical performance, GCNs have become a foundational model in many graph learning tasks and have inspired numerous subsequent GNN architectures.

2.2.3 Graph Representation Learning with GNNs

One of the primary goals of graph neural networks is to learn informative representations of nodes, edges, or entire graphs. By iteratively aggregating structural and attribute information from neighboring nodes, GNNs are able to transform raw

graph data into low-dimensional vector representations that can be used for various downstream machine learning tasks.

At the node level, GNNs are widely used for node classification tasks. In this setting, the objective is to predict the label of each node based on both its attributes and its structural position in the graph. After K layers of message passing, the final representation of node v can be written as

$$z_v = h_v^{(K)} \tag{2.8}$$

where $h_v^{(K)}$ denotes the embedding of node v after K propagation layers. These learned node embeddings can then be used as input features for classification or regression models.

Beyond node-level prediction, GNNs can also be applied to link prediction tasks. Link prediction aims to estimate the likelihood that an edge exists between two nodes or that a new connection may appear in the future. This task is particularly useful in applications such as recommendation systems, social network analysis, and biological interaction prediction.

GNNs can also generate representations for entire graphs. In graph-level tasks, information from all node embeddings is aggregated through a readout operation to produce a graph representation

$$z_G = \text{READOUT}(\{h_v^{(K)} | v \in V\}) \tag{2.9}$$

where the READOUT function may correspond to operations such as summation, averaging, or max pooling across node embeddings. The resulting graph representation can then be used for tasks such as graph classification or graph regression.

Through these different learning paradigms, graph neural networks provide a flexible framework for extracting meaningful patterns from graph-structured data across multiple levels of granularity.

2.2.4 Applications of Graph Neural Networks

Due to their ability to jointly model node attributes and relational structures, graph neural networks have been successfully applied across a wide range of domains where data naturally exhibit graph structures. By capturing complex dependencies between entities, GNN-based models are able to uncover patterns that may be difficult to detect using traditional machine learning approaches.

In financial systems, many types of data can be naturally represented as networks. For example, transaction networks can be constructed where nodes represent accounts and edges represent financial transfers. Modeling such structures enables the identification of suspicious interaction patterns that may indicate fraudulent activities. GNN-based approaches have therefore been widely explored in tasks such as fraud detection, anti-money laundering analysis, and risk propagation modeling. By incorporating relational information among entities, these models are able to improve detection accuracy and better capture coordinated fraudulent behaviors.

Graph neural networks have also demonstrated strong potential in biomedical and healthcare research. Many biological systems can be represented as networks, including gene regulatory networks, protein–protein interaction networks, and disease association networks. In these settings, GNNs can integrate biological features with interaction structures to better understand complex biological mechanisms. Applications include disease gene prediction, drug discovery, biological pathway analysis, and modeling disease progression. By leveraging both molecular features and biological interactions, graph-based learning methods can provide valuable insights into the underlying mechanisms of complex diseases.

Beyond finance and biomedical applications, GNNs have also been widely used in recommendation systems, knowledge graph reasoning, transportation networks, and social network analysis. These diverse applications highlight the flexibility of graph neural networks in modeling relational data. As graph-structured datasets continue to grow in scale and complexity, GNN-based approaches are expected to

play an increasingly important role in extracting meaningful knowledge from interconnected systems.

2.3 Fairness in Machine Learning and Graph Learning

As machine learning systems are increasingly deployed in real-world decision-making processes, concerns regarding fairness and bias have attracted significant attention from both researchers and practitioners. In many high-stakes applications, such as credit approval, hiring decisions, and medical diagnosis, machine learning models may influence outcomes that directly affect individuals or social groups. If these systems learn biased patterns from historical data, they may produce predictions that systematically disadvantage certain groups of people. Consequently, ensuring fairness in machine learning models has become an important research topic aimed at mitigating discrimination and promoting equitable decision-making.

Bias in machine learning systems can arise from multiple sources. One common cause is historical bias embedded in training data, where past social or institutional inequalities are reflected in the collected datasets. Another source is sampling bias, where certain groups are underrepresented or overrepresented in the data, leading to imbalanced learning outcomes. In addition, algorithmic bias may occur when model architectures or optimization procedures amplify existing disparities in the data. These issues have motivated the development of fairness-aware machine learning approaches that aim to reduce discriminatory behaviors while maintaining predictive performance.

2.3.1 Fairness in Machine Learning

To address fairness concerns, a wide range of techniques have been proposed to incorporate fairness considerations into machine learning pipelines. These approaches

can generally be categorized into three groups based on the stage at which fairness interventions are applied.

The first category includes pre-processing methods, which attempt to mitigate bias by modifying the training data before the learning process begins. Typical strategies involve reweighting samples, removing sensitive attributes, or transforming the dataset so that different demographic groups are represented more equally. By improving the fairness of the training data itself, these methods aim to reduce the likelihood that models will learn discriminatory patterns.

The second category consists of in-processing approaches, which integrate fairness constraints directly into the model training process. In these methods, fairness objectives are incorporated into the loss function or optimization procedure, encouraging the model to balance predictive accuracy with fairness considerations during learning. This approach allows fairness to be enforced while the model parameters are being estimated.

The third category includes post-processing techniques, which adjust the model outputs after training has been completed. In this case, fairness is achieved by modifying prediction thresholds or recalibrating the predicted probabilities for different groups. Although these methods do not change the internal structure of the trained model, they provide a flexible way to enforce fairness constraints at the decision-making stage.

2.3.2 Types of Fairness: Group Fairness and Individual Fairness

Within the fairness literature, two major concepts are commonly discussed: group fairness and individual fairness. These definitions provide different perspectives on what it means for a machine learning system to behave fairly.

Group fairness focuses on ensuring that different demographic groups receive similar outcomes from the model. In many applications, individuals can be divided

into groups based on sensitive attributes such as gender, race, or age. A common formulation of group fairness requires that the probability of receiving a positive prediction should be similar across these groups. For example, the concept of statistical parity requires that

$$P(\hat{Y} = 1 \mid A = 0) = P(\hat{Y} = 1 \mid A = 1) \tag{2.10}$$

where A represents a sensitive attribute and $\hat{Y}$ denotes the predicted outcome. This definition aims to ensure that the model does not systematically favor one group over another.

In contrast, individual fairness emphasizes the treatment of similar individuals. The central idea is that if two individuals are similar with respect to relevant attributes, they should receive similar predictions from the model. Formally, individual fairness can be expressed as

$$d(f(x_i), f(x_j)) \leq \epsilon \tag{2.11}$$

when the distance between two individuals x_i and x_j is small according to an appropriate similarity metric. This formulation encourages consistent treatment across individuals with comparable characteristics.

Although both definitions aim to promote fairness, they focus on different aspects of equity and may not always be satisfied simultaneously. As a result, balancing group fairness and individual fairness has become an important challenge in fairness-aware machine learning research.

2.3.3 Fairness Challenges in Graph Neural Networks

While fairness has been extensively studied in traditional machine learning settings, addressing fairness in graph-based learning introduces additional challenges. Graph neural networks operate on relational data where nodes are interconnected through edges, meaning that the prediction for one node may depend on information from

many neighboring nodes. This relational dependency can lead to the propagation of biases across the network.

One important factor is the structural dependency inherent in graph data. Since node representations are updated by aggregating information from neighbors, biased information associated with certain nodes may influence the representations of many other nodes through message passing. As a result, unfair patterns present in a small portion of the graph may spread across the network during training.

Another challenge arises from the phenomenon of homophily, where nodes with similar attributes tend to connect with one another. While homophily can be beneficial for many prediction tasks, it may also reinforce existing biases if sensitive attributes correlate with the graph structure. In such cases, graph learning algorithms may unintentionally amplify discriminatory patterns that already exist in the network.

Furthermore, fairness definitions originally developed for traditional machine learning models may not directly translate to graph learning scenarios. The presence of complex dependencies between nodes complicates the interpretation of fairness metrics and makes it more difficult to design algorithms that simultaneously maintain predictive accuracy and fairness.

These challenges highlight the need for fairness-aware approaches specifically designed for graph neural networks. Developing methods that balance fairness objectives while preserving the expressive power of graph-based models remains an important direction in the study of fair graph learning.

2.4 Machine Learning in Biomedical Data Analysis

Recent advances in biomedical technologies have led to the rapid growth of large-scale biological datasets. High-throughput experimental techniques now allow researchers to measure molecular activities at an unprecedented scale, generating data that describe complex biological processes across different levels of cellular organi-

zation. As a result, machine learning has become an important tool in biomedical research, enabling the discovery of patterns and relationships that may not be easily identified through traditional statistical approaches.

Machine learning methods have been widely applied to a variety of biomedical tasks, including disease diagnosis, biomarker discovery, drug development, and personalized medicine. By learning patterns from large and complex datasets, machine learning models can assist researchers in identifying disease-associated genes, understanding biological mechanisms, and predicting disease progression. These capabilities make machine learning particularly valuable in modern biomedical data analysis, where the volume and complexity of available data continue to increase.

2.4.1 Machine Learning in Genomic and Transcriptomic Data

Among the many types of biomedical data, genomic and transcriptomic datasets have become especially important for studying molecular mechanisms of diseases. Genomic data describe the structure and variation of DNA sequences, while transcriptomic data measure gene expression levels that reflect the functional activity of genes within cells. Advances in sequencing technologies have enabled the generation of large-scale omics datasets, including genomics, transcriptomics, proteomics, and metabolomics. These complementary data sources have motivated the development of multi-omics approaches that aim to integrate multiple biological data modalities to better understand complex biological systems.

Among these different data types, transcriptomic data derived from RNA sequencing (RNA-seq) has become one of the most widely used resources for analyzing gene expression patterns. RNA-seq technologies allow researchers to measure the expression levels of thousands of genes simultaneously, providing detailed insights into cellular functions and biological processes. By analyzing gene expression profiles across different biological conditions, researchers can identify genes that are associated with disease progression, uncover potential biomarkers, and explore molecular

pathways involved in complex diseases.

Machine learning techniques play an important role in extracting meaningful information from transcriptomic datasets. By combining statistical learning methods with biological data analysis, machine learning models can identify patterns in high-dimensional gene expression data and predict disease-related outcomes. Applications of machine learning in transcriptomic analysis include disease classification, biomarker discovery, and the identification of genes associated with specific pathological conditions. These approaches provide valuable tools for understanding the molecular basis of diseases and for developing new therapeutic strategies.

2.4.2 Challenges in Biomedical Data Analysis

Despite the potential of machine learning in biomedical research, analyzing biomedical datasets presents several significant challenges. One of the most prominent issues is the high dimensionality of biological data. For example, transcriptomic datasets often contain expression measurements for thousands of genes, while the number of available samples may be relatively limited. This imbalance between feature dimensionality and sample size can make it difficult for machine learning models to generalize effectively.

Another challenge is the limited availability of labeled data. In many biomedical studies, collecting large datasets is difficult due to experimental costs, ethical considerations, and the complexity of biological experiments. As a result, many datasets used in biomedical research contain relatively small sample sizes, which may increase the risk of overfitting in machine learning models.

In addition, biological systems are inherently complex and involve intricate interactions among genes, proteins, and cellular pathways. These interactions can create nonlinear relationships that are difficult to capture using simple analytical methods. Understanding such complex biological dependencies requires machine learning approaches that are capable of modeling high-dimensional and heteroge-

neous data.

These challenges highlight the importance of developing effective machine learning frameworks for biomedical data analysis. By designing methods that can handle high-dimensional data, limited samples, and complex biological relationships, machine learning models can help uncover meaningful biological insights and support the study of disease mechanisms.

2.4.3 Graph Learning for Biomedical Data Analysis

Many biological systems can naturally be represented as networks, where nodes correspond to biological entities and edges describe interactions between them. Examples include gene regulatory networks, protein–protein interaction networks, metabolic networks, and disease–gene association networks. These network representations provide an intuitive way to model the complex relationships that exist among molecular components in biological systems.

Traditional machine learning approaches often treat samples as independent observations and primarily rely on feature-based representations. However, in biological systems, the relationships between entities are often as important as the attributes of the entities themselves. Graph-based learning methods provide a natural framework for capturing these interactions by modeling both node features and network structures simultaneously. By leveraging relational information embedded in biological networks, graph learning algorithms can uncover patterns that may not be detectable through conventional analysis methods.

Graph neural networks have therefore attracted increasing interest in biomedical research. They have been applied to tasks such as gene function prediction, disease gene identification, drug discovery, and biological pathway analysis. By integrating molecular features with network topology, graph-based learning approaches offer promising opportunities for understanding complex biological mechanisms and improving the analysis of large-scale biomedical datasets.

Taken together, the literature reviewed in this chapter demonstrates the remarkable progress of graph-based machine learning across financial and biomedical applications. Nevertheless, most existing studies primarily emphasize improvements in graph neural network architectures or learning algorithms, while comparatively less attention has been devoted to the practical challenges associated with applying graph-based learning in high-stakes real-world settings. These challenges include heterogeneous and noisy data, fairness requirements, limited supervision, domain-specific interpretability, and dynamic biological processes. Motivated by these research gaps, this thesis investigates how graph-based machine learning systems can be systematically designed to address practical constraints arising in financial and biomedical applications. The following chapters investigate this overarching objective from three complementary perspectives: graph representation learning for financial fraud detection, fairness-aware graph learning, and graph-based machine learning for biomedical intelligence.

Chapter 3

Heterogeneous Graph Learning for Credit Card Fraud Detection

The global economic losses caused by credit card fraud are enormous and continuously increasing. Effective and accurate fraud detection has become a crucial task in recent years. Prior approaches can achieve good detection performance under certain conditions. However, these existing methods lack the robustness and scalability to deal with real-world credit card transactional datasets containing a large number of missing values. In this chapter, we propose a Feature Importance-based Weighted Graph Neural Network (FIW-GNN) as an effective, stable, and practical solution for credit card fraud detection. First, we propose a method to construct a heterogeneous graph designed for credit card transactional datasets. Next, based on the architecture of the relational graph convolutional network, a feature importance-based method is employed to assign edge weights. Finally, we evaluate the effectiveness of FIW-GNN on two benchmark datasets. The experimental results demonstrate that FIW-GNN outperforms the state-of-the-art baselines in all selected evaluation metrics.

3.1 Introduction

Credit card payment has been widely popularized in the past few decades, and related fraud crimes have continuously intensified. Although the proportion of fraudulent activities is small, the economic losses it can cause are unprecedentedly huge. Payment card fraud resulted in a total of $499.5 million in financial losses nationwide in Australia in the 12 months to June 30, 2022, according to the latest statistics from the Australian Payments Network (Network, 2022). Moreover, based on the latest Nilson report (Report, 2021), credit card fraud caused $28.58 billion in economic losses worldwide in 2020, and this amount is projected to be $49.32 billion in 2030. Therefore, effective and accurate credit card fraud detection has become an urgent and challenging task.

Data mining and machine learning are widely used methods in credit card fraud detection (Awoyemi et al., 2017). In practice, classic machine learning models such as support vector machine (SVM) (Sahin & Duman, 2010), decision tree (DT) (Husejinovic, 2020), and logistic regression (LR) (Itoo et al., 2021) have been applied and performed well in fraud detection. However, such methods are unstable when dealing with different datasets and application scenarios, and perform poorly when dealing with real-world transactional datasets with higher dimensions and more noise. Deep learning is another popular approach to tackling credit card fraud detection. Some studies are using convolutional neural networks (CNN) (Fu et al., 2016) and recurrent neural network (RNN) (Roy et al., 2018) based models, and it has been demonstrated that these deep learning models have more robust effectiveness and stability than traditional machine learning models. However, such models often emphasize the utilization of feature selection and feature engineering techniques in the data preprocessing stage, which leads to the loss of a large amount of raw data. Graph neural network (GNN) has attracted much attention recently and has made breakthroughs in many financial applications (J. Wang et al., 2021). Some studies have used GNN-based methods for credit card fraud detection, and

progress has been achieved in detection performance (G. Liu et al., 2021) (Jing et al., 2021). However, the existing methods do not have a high utilization rate of the original data, and the processing methods for missing values and edge definition lack interpretability.

To address to the aforementioned drawbacks, we propose a novel GNN-based framework for credit card fraud detection. We design a heterogeneous graph that can fully use the effective information of the original dataset without being affected by missing values, named as Credit-card HEtero Transactional graph (CHET-graph). We also propose an edge weight assigning approach based on feature importance using the Chi-squared statistic method. Finally, a combination of CHET-graph, feature importance-based edge weights, and relational graph convolutional network (RGCN) architecture, FIW-GNN, is proposed to obtain a better detection performance in real-world credit card transactional datasets. To validate the effectiveness of FIW-GNN, we conduct empirical experiments compared to several baseline models, using a real-world dataset sourced from Vesta's e-commerce transactions and a simulated credit card dataset under a public simulator.

The main contributions of our study are summarized as follows:

- We propose an innovative way of building a new heterogeneous graph for real-world credit card transactional datasets, namely CHET-graph, which can overcome the problem of a large number of missing values and maximize the use of raw data.

- We design a novel method for assigning weights in heterogeneous graphs based on feature importance to increase the interpretability of graphs.

- We propose a novel GNN-based model based on CHET-graph, feature importance-based weights assignment, and RGCN architecture, namely FIW-GNN, to achieve better detection performance in credit card fraud.

- Experimental results on two benchmark credit card fraud datasets demonstrate

the effectiveness of the proposed method.

The remainder of this chapter is organized as follows. Section 3.2 presents a literature review of credit card fraud detection on machine learning, deep learning, and GNN approaches. Section 3.3 illustrates our proposed FIW-GNN framework. Section 3.4 introduces the experiments from the dataset description, compared methods, and evaluation metrics. Section 3.5 presents and discusses the experimental results, and Section 3.6 summarizes the chapter's conclusions and future work.

3.2 Related Work

Research on credit card fraud detection has been carried out for many years. The technical methods are constantly and iteratively updated, and the detection performance is also continuously improved. We summarize the related work in three main approaches: 1) classic machine learning approach, 2) deep learning approach, and 3) graph neural network approach.

3.2.1 Classic Machine Learning Approach

In recent years, classic machine learning-based methods have been widely used in credit card fraud detection. For example, Logistic regression (LR) has been used as a predictive classifier for improved detection of credit card fraud. Hussein *et al.* (Hussein et al., 2021) proposed a stacking ensemble algorithm by combining LR with the fuzzy-rough nearest neighbor (FRNN) and sequential minimal optimization (SMO), which reached a detection rate of 84.90% on an Australian credit approval dataset. Another comparative analysis study by Itoo *et al.* (Itoo et al., 2021) deploys three classic machine learning methods, namely LR, Naïve Bayes, and K-nearest neighbor (KNN), for a transactional fraud detection task. They used under-sampling techniques and demonstrated the LR model outperforms the others on the selected measurements, such as F-measure and area under a curve (AUC). Random

Forest (RF) was also used to detect fraudulent transactions and obtained higher accuracy than other machine learning algorithms. The studies (Jemima Jebaseeli et al., 2021; Kumar et al., 2019) showed that RF relies on the quality and relevance of multiple decision trees and works well with small amounts of data. In addition, a hybrid framework using AdaBoost and majority voting methods was deployed by Randhawa *et al.* (Randhawa et al., 2018) in detecting fraud cases in credit card transactions. The robustness and reliability of the AdaBoost model have been demonstrated under a real-world credit card dataset with additional noise.

However, the limitations of classic machine learning methods in real-world credit card fraud detection are also evident. Firstly, there is still a lot of room for improvement in the detection performance of this type of methods. Also, there is a lack of stability. The performance of the same machine learning model in different datasets is often unstable. Secondly, the evaluation metrics selected in the machine learning approach are not convincing enough. Due to the highly imbalanced nature of credit card transaction datasets, evaluation metrics like accuracy are inappropriate. Thirdly, these classic machine learning methods do not perform well when facing massive data with high-dimensional features, thereby emphasizing feature engineering and selection techniques (Singh & Jain, 2019) (Lucas et al., 2020). Consequently, classic ML approaches are not practical for large real-world credit card transaction data.

3.2.2 Deep Learning Approach

The machine learning methods showed both good performance and shortcomings in detecting credit card fraud. As a branch that has continuously made breakthroughs in machine learning, deep learning methods promissed higher detection performance and stability. Some recent studies have adopted deep learning methods based on convolutional neural networks (CNN). For example, Fu *et al.* (Fu et al., 2016) proposed a CNN-based credit card fraud pattern mining framework. The CNN

model identifies latent patterns for each sample on the feature matrix transformed by transactional data and demonstrates its superior performance in fraud detection. CCFD-Net, proposed in (X. Liu et al., 2021), achieved better detection performance on a credit card transaction dataset from Vesta company compared to seven other machine learning models. It is a hybrid CNN-based architecture combining the 1D-Conv technique and the residual neural network (ResNet), demonstrating sufficient robustness.

On the other hand, based on the sequential characteristics of credit card transaction data, recurrent neural network (RNN) has also been applied to fraud detection in many studies. The studies conducted by Roy *et al.* (Roy et al., 2018) and Jurgovsky *et al.* (Jurgovsky et al., 2018) both employed Long Short-Term Memory (LSTM) networks as sequence learners in solving common fraud detection problems. The results prove that sequence learners have better efficacy in transactional fraud detection than traditional static learners such as RF. The model utilized in (Najadat et al., 2020) integrates bidirectional LSTM and Gated recurrent unit (GRU), which also obtains better fraud detection performance than six machine learning classifiers.

However, compared with the classic machine learning approaches, the research in the deep learning approach is not comprehensive enough. Moreover, the deep learning models used in these studies are mostly transplanted from other application fields, which lack improvements and innovations designed for specific credit card transaction scenarios. They cannot fully meet the characteristics of credit card datasets. Like the classic machine learning approaches, deep learning architectures such as HOBA (X. Zhang, Han, et al., 2021) still rely heavily on feature engineering and selection techniques.

3.2.3 Graph Neural Network Approach

We can convert transactional data into graph representations through network embedding, which aims at inferring a low-dimensional vectorized representation for each node. Recently, graph neural network (GNN) methods, such as GCN (Welling & Kipf, 2017) and GraphSAGE (Hamilton et al., 2017), have been applied to detect fraudulent credit card transactions since GNN is capable of handling graph data and high-order cross-features.

Liu *et al.* (G. Liu et al., 2021) designed a weighted homogeneous graph that takes each transaction record as a node and defines different logical propositions as the conditions for generating edges. Based on the homogeneous graph constructed, a GNN model combining the GraphSAGE algorithm, a novel sampling policy, and an attention-based aggregation mechanism was utilized for detecting fraudulent transaction behaviour. Another GNN-based model which focuses on credit card fraud detection with few samples is proposed in (Jing et al., 2021). It builds an adjacency matrix based on learnable parameters, performs message passing according to the similarity of features, and calls the GCN layer to extract node features. Additionally, in (Cheng et al., 2020), a spatial-temporal attention-based graph network (STAGN) is proposed and demonstrated with better fraud detection performance in both AUC and precision-recall curves. The STAGN model learns representations from the global transaction graph, which is constructed by transforming raw transaction records into spatial-temporal-based feature tensors. Some other works, such as PC-GNN (Y. Liu et al., 2021), CARE-GNN (Dou et al., 2020), Know-GNN (Rao et al., 2021), and Lambda architecture GNN (Lu et al., 2021), also reflect the excellence of the GNN-based methods in improving the performance of fraud detection, which has significant advantages compared with classic machine learning and deep learning approaches.

However, the GNN methods currently used in credit card fraud detection also have apparent flaws. Firstly, these methods do not fully consider all raw features in

constructing transaction graphs. Most of these methods use homogeneous graphs, which cause certain limitations. Secondly, some methods use expert knowledge to define the composition of edges or the assignment of weights, such as logical propositions mentioned in (G. Liu et al., 2021). These types of assumptions are less convincing and less scalable. There is a lack of uniform definition standards for credit card transaction datasets in different scenarios, resulting in a lack of robustness.

3.3 Methodology

In this section, we first present the method to construct a novel heterogeneous graph designed for the credit card transactional dataset: **C**redit-card **HE**tero **T**ransactional **graph** (**CHET-graph**). Then we introduce an overview of Relational Graph Convolution Network (RGCN) (Schlichtkrull et al., 2018). After that, we describe the process of assigning edge weights based on categorical feature importance by the Chi-squared method and mutual information method, respectively. Eventually, we present the final framework, **F**eature **I**mportance-based **W**eighted **G**raph **N**eural **N**etwork, namely **FIW-GNN**.

3.3.1 CHET-graph

The advanced deep learning methods and the existing GNN-based methods face one problem in the data preprocessing stage. When dealing with a credit card transaction dataset with a large number of missing values, these methods either use a filling method that lacks explanation or delete a large number of raw features. In addition, most of the existing GNN methods for credit card fraud detection use homogeneous graphs, defining transaction records as nodes in the graph. The edges and corresponding weights between different nodes are defined based on the attributes of transaction instances. It leads to the following three problems: (1) A large number of original features, especially the categorical textual features, cannot be exploited. (2) The definition and weight of edges often rely on expert knowledge or

subjective assumptions. (3) The applicability is not ideal, and the graph definition method on one dataset cannot be conveniently applied to other datasets. Credit card transaction datasets in the real world often have many missing values and differences in feature attributes. The discarded raw features in the above methods contain valuable original information.

In this chapter, we aim to find an objective way to define a graph for credit card transaction datasets that can comprehensively contain all original feature information and is scalable for different scenarios. Therefore, we propose the CHET-graph.

Definition 1. Transaction Record. A transaction record r is a tuple of attributes in a payment instance, i.e. $r := \{a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n\}$, where a_i denotes the i-th node attribute, which is used as node feature of target-type in the CHET-graph, and b_j denotes the j-th edge attribute which is used to represent edge types and define nodes other than target-types.

Definition 2. Transaction Slice. Given a transaction record r, its transaction slice r_t is the set of attributes that reflect each transaction instance, containing information such as transaction amount, time interval, match situation, etc.: $r_t := \{a_1, a_2, \cdots, a_m\}$.

Definition 3. Identity Slice. Given a transaction record r, its identity slice r_i is the set of attributes representing the identity information within the transaction instance, such as card information, address, email domain, device information, etc.: $r_i := \{b_1, b_2, \cdots, b_n\}$. The identity slices reflect the relations between transactions.

Next, we introduce the construction process of a CHET-graph from a credit card transaction dataset. An example of the CHET-graph is shown in Fig. 3.1.

(1) Feature segmentation. We first divide all features of transaction instances into transaction slices and identity slices. The features in transaction slices are treated as numerical. If we encounter a categorical feature, we use the one-hot encoding method to convert it into two or more numerical features. On the other hand, the features classified into identity slices are all treated as categorical, in which some will contain textual information but do not require any conversion.

(2) Node extraction and generation. First, we create a targe-type node for each transaction record and use the transaction slice of this record as the features of the node. Next, we create an edge-type node for each categorical feature in the identity slice of this record. The number of features in the identity slice determines the number of edge-type nodes in the CHET-graph. For example, suppose there is an identity slice containing three categorical features (*Card-type*, *Device-type*, *Email-domain*), and each categorical feature has three distinct categorical values, as shown in Fig. 3.1. Then, there will be three different categories of edge-type nodes in the CHET-graph, corresponding to "Card-type", "Device-type", and "Email-domain", respectively. Each category has three nodes, and there are nine edge-type nodes in total. As shown in Fig. 3.1, "Device-type" is a categorical feature in the identity slice, which has three categorical values, "mobile", "desktop" and "terminal". It contributes three nodes to the final CHET-graph. Therefore, nodes in the CHET-graph can be divided into two categories: the target-type nodes corresponding to the transaction instances and the edge-type nodes corresponding to all categories of the categorical features in the identity slice.

(3) Edge generation. The edges in the CHET-graph will be generated between target-type nodes and different kinds of edge-type nodes. We generate corresponding edges based on the values in the features of the identity slice of a target-type node. For example, if the "Device-type" feature value of a certain transaction record is "mobile", then the target-type node corresponding to this transaction record will generate an edge with the edge-type node "mobile". Then, all the edges generated between target-type nodes and "Device-type" nodes consist an edge set corresponding to the relation of "target to Device-type". The process is the same with the other categories of edge-type nodes. The orange, blue and green edges in Fig. 3.1 corresponds to three edge sets of relation "target to Card-type", "target to Device-type", and "target to Email-domain", respectively. Here, the absence of edges due to missing values does not affect the construction of the graph. Therefore, CHET-graph can fully use all the information in the raw dataset while handling the

problem of enormous missing values.

(4) Edge weight assignment. For the edges generated between the target-type nodes and edge-type nodes, it is simple to default the weight of all edges to 1. However, such an approach lacks interpretability. Here, we propose a feature importance-based approach to assign weights to edges in a CHET-graph, which will be elaborated on in later sections. For the example above, the edges generated by the target-type nodes and the edge-type nodes belonging to categories "Card-type", "Device-type", and "Email-domain" will obtain different weight values according to the feature importance.

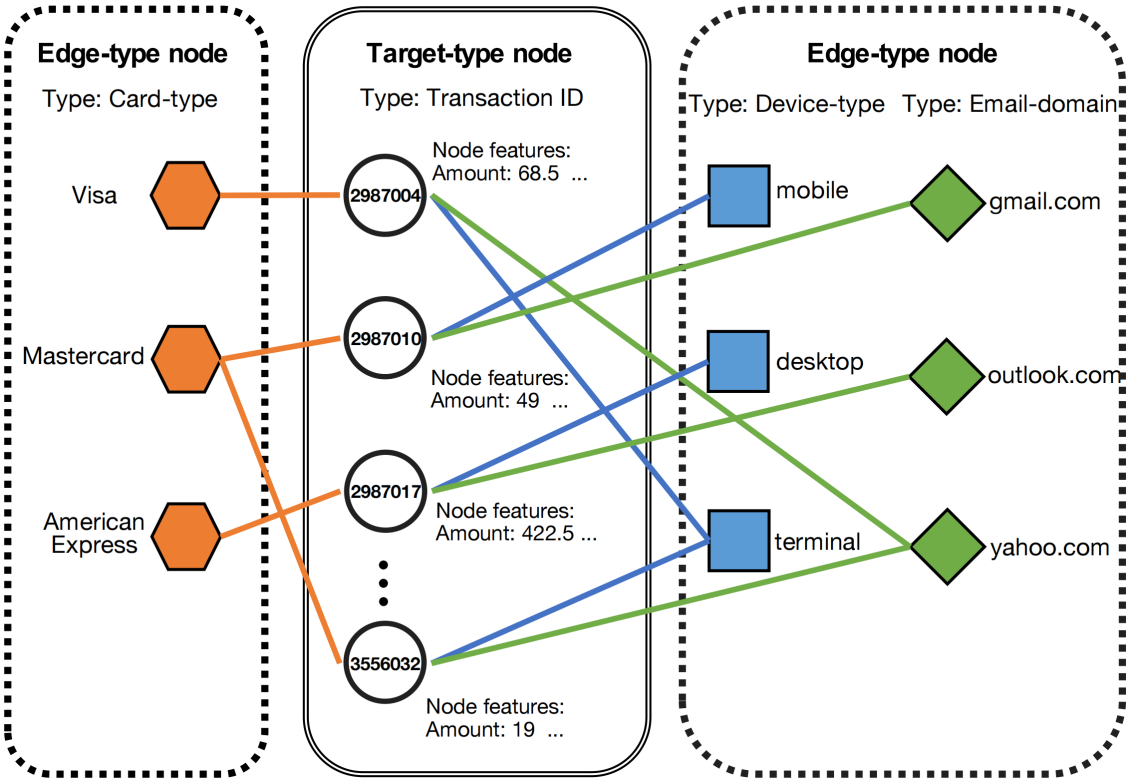


Figure 3.1: An example of the CHET-graph.

In summary, CHET-graph is a heterogeneous graph containing different types of nodes and weighted edges. We present the definition of the CHET-graph as follows:

Definition 4. CHET-graph. Let $V_t = \{v_{t_1}, v_{t_2}, \cdots, v_{t_N}\}$ be the set of target-type nodes and $V_e = \{v_{e_1}, v_{e_2}, \cdots, v_{e_M}\}$ the set of edge-type nodes. Denote $V = V_t \cup V_e$ as the node set. For each of R relations including both in canonical and inverse

directions, E_r is the set of node pairs and A_r is the corresponding adjacency matrix under the relation r ($1 \leq r \leq R$). Denote $E = E_1 \cup \cdots \cup E_R$ and $A = \{A_1, ..., A_R\}$, then the heterogeneous graph $G = \{V, E, A\}$ is called a *CHET-graph*. Particularly we use X of size $t_N \times d$ to represent the feature matrix over all the target-type nodes V_t, and relevant to the adjacency matrices, we use $W = \{W_1, W_2, ..., W_R\}$ to denote the weight matrices for R relations.

3.3.2 RGCN

RGCN is a model for processing large-scale relational data constructed on the basis of GCN (Welling & Kipf, 2017). It defines the following simple propagation model for calculating the forward-pass update of a node in a relational multi-graph:

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right) \tag{3.1}$$

where h_i^l is the hidden state of node v_i in the l-th layer of RGCN. σ is an activation function such as ReLU. $\mathcal{N}_i^r$ denotes the set of neighbor indices of node i under relation $r \in \mathcal{R}$. $c_{i,r}$ is a problem-specific normalization constant that can be set in advance.

The neural network layers can be stacked to allow for dependencies across several relational steps. Each node in the graph will be evaluated and updated in parallel. Two forms of relations exist in a CHET-graph from the perspective of edge generation. One is between the target-type node and the edge-type node. The other is the self-connected edges generated by each target-type node itself, which is known as the relation-specific transformation in RGCN. The purpose is to ensure that the representation of a node at layer $l + 1$ can also be notified by the corresponding representation at layer l. The feature vectors of neighboring nodes are gathered and transformed for each relation type individually. Then, the transformed representations are accumulated by normalized sum and sent to a RELU function. Fig. 3.2

shows the process of computing the update for a single node in a CHET-graph under the RGCN model.

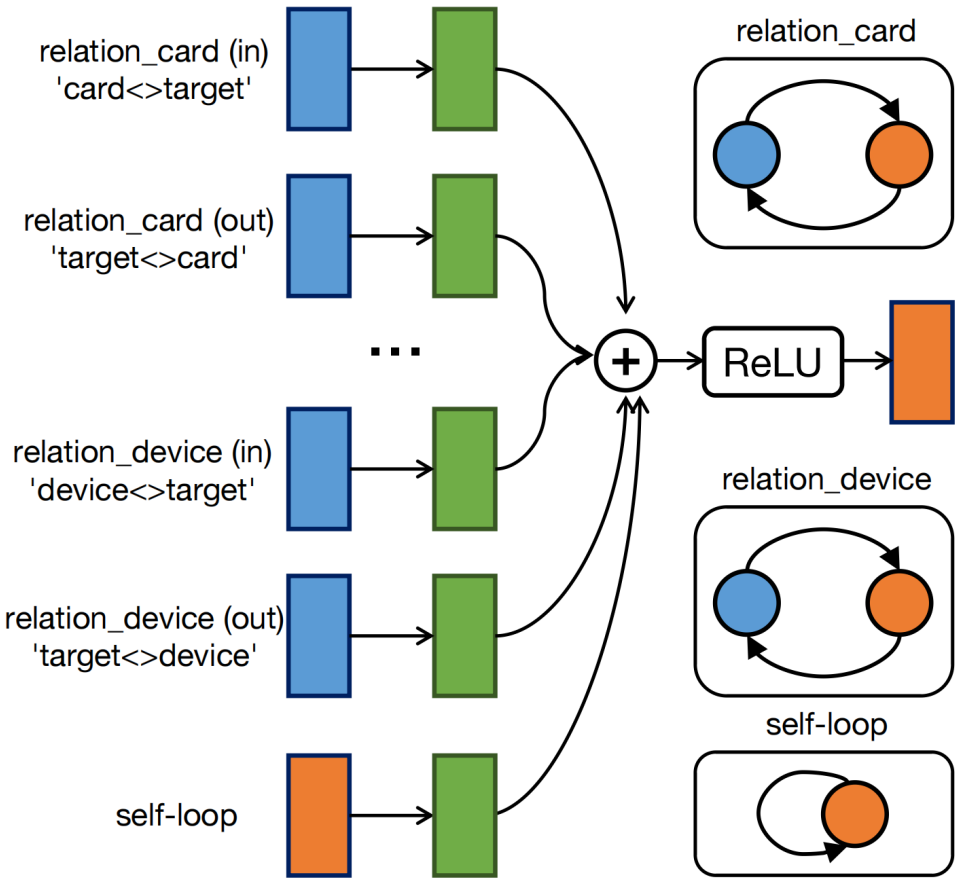


Figure 3.2: Computation process for a node in CHET-graph using RGCN model.

To complete the classification task of each target-type node in the CHET-graph, we minimize the cross-entropy loss on all labeled nodes and apply a softmax function on the output of the last layer.

3.3.3 Feature Importance-based Edge Weights

RGCN aims to train heterogeneous graphs with multiple edge types while does not consider edge features. It defaults edge only has different types and fits well with the CHET-graph's characteristic of "homogeneous nodes, heterogeneous edges". In most methods of solving financial problems using the RGCN architecture or

related Python libraries (Data61, 2018), the weights of different types of edges are not taken into account. In other words, the weights of all edges are equal to 1. Although this approach simplifies the definition of such heterogeneous graphs while ensuring good training results, it will lead to a lack of interpretability in credit card fraud detection. Therefore, we propose a method for assigning edge weights based on feature importance, which enables different types of edges in the CHET-graph to obtain weights that match their feature importance. It can increase the interpretability of the model and, in practice, further improve the performance of fraud detection.

The type of edges in the CHET-graph is determined by the features in the identity slice. Most of the features in these identity slices contain textual information, which will be treated as categorical features. Here, we use two approaches to extract the importance values of these categorical features: the Chi-squared statistic method and the mutual information statistic method.

(1) Chi-squared statistical method. The Pearson's Chi-squared test is used to quantify the independence of pairs of categorical variables. Pairs of categorical variables can be summarized in a contingency table, reflecting their relation. We calculate the expected and observed values based on the contingency tables. Afterwards, the Chi-square values of each categorical feature, which will be used as the original feature importance scores, can be calculated using:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \tag{3.2}$$

where the summation is taken over the contingency table, O is the count of observations, and E is the calculated expected value.

(2) Mutual information statistical method. Mutual information is an application of information gain and measures the reduction in entropy under the condition of the target value. Particularly, in our case, the conditional entropy of

discrete random variables X given Y is defined as:

$$H[X \mid Y] = -\sum_{x}\sum_{y} p(x,y)\log_2 p(x \mid y) \tag{3.3}$$

where the summations are taken over all the categorical values of X and Y, respectively. $H[X \mid Y]$ can be interpreted as the expected number of additional bits required to encode X given that the value of Y is already known.

Then, the Mutual Information score can be calculated using (3.4).

$$MI[X,Y] = H[X] - H[X \mid Y] = H[Y] - H[Y \mid X] \tag{3.4}$$

The MI scores range from 0 to ∞, which will be normalized and used as the feature importance values. The higher the value, the higher the weight of the corresponding type of edges in the CHET-graph.

3.3.4 FIW-GNN

Based on the construction of the CHET-graph, the architecture of RGCN, and the logic of edge weights based on feature importance, we proposed the Feature Importance-based Weighted Graph Neural Network (FIW-GNN). Fig. 3.3 shows the framework of our proposed FIW-GNN method.

After obtaining the original credit card transaction dataset, we divide it into transaction and identity slices, and generate corresponding nodes and edges. Then, we use a method based on feature importance to assign edge weights and construct a CHET-graph. Afterward, the CHET-graph is used as the input of RGCN. Each node is computed and updated in parallel, and is output through a softmax activation in the last layer. Eventually, we receive the prediction results for evaluation in the later stage.

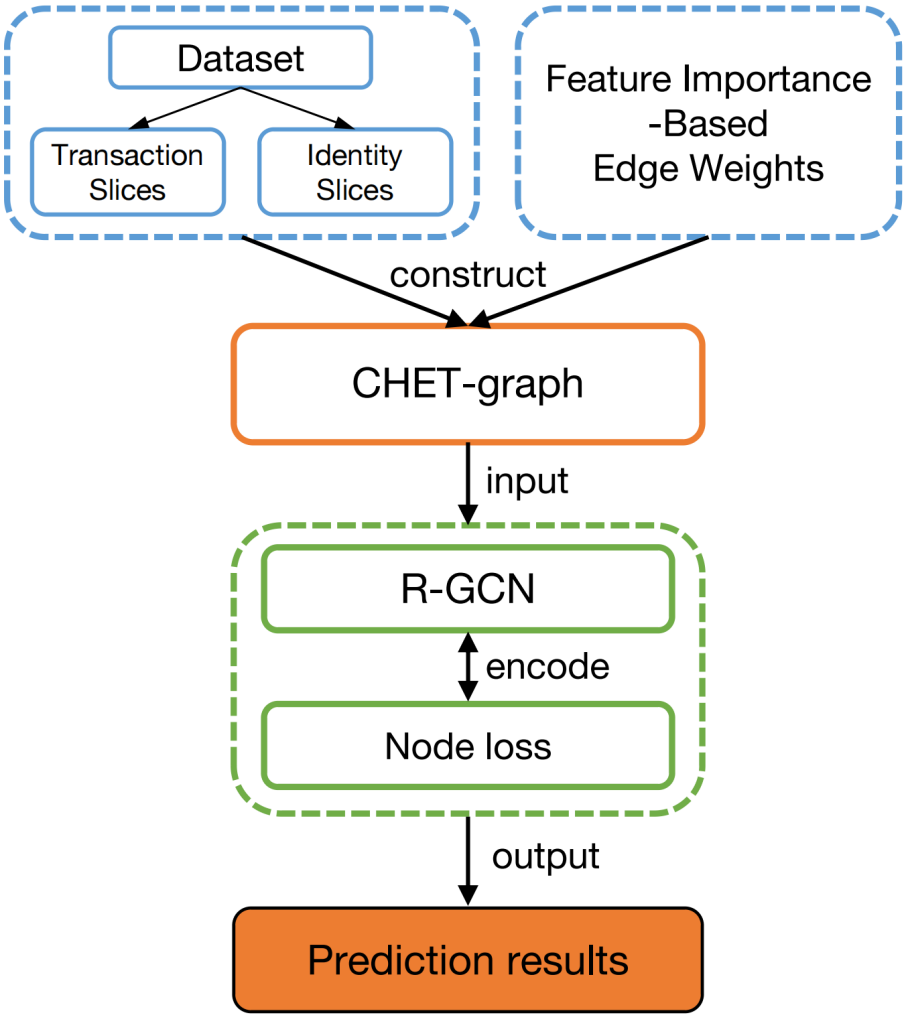


Figure 3.3: The framework of proposed FIW-GNN method.

3.4 Experiments

3.4.1 Datasets Description

This study uses one real-world credit card transaction dataset and one simulated credit card transaction dataset. The Vesta dataset[1] is collected from an American corporation's e-commerce transactions in 2019. The Sparkov dataset[2] is generated using the Sparkov Data Generation tool by Brandon Harris between 2019 and 2020.

[1]https://www.kaggle.com/competitions/ieee-fraud-detection/data
[2]https://www.kaggle.com/datasets/kartik2112/fraud-detection

Both datasets are heavily skewed and contain a large number of missing values. The detailed statistics of these two datasets are shown in Table 3.1.

Table 3.1: Statistics of Datasets.

Composition	**Vesta Dataset**	**Sparkov Dataset**
Samples	472,432	1,296,675
Dimensions of features	432	22
Fraud instances (minority class)	16,599	7,506
Legit instances (majority class)	455,833	1,289,169
Fraud ratio	3.5%	0.6%
Missing value	92,538,150	12,837,082
Missing value ratio	45.34%	45.01%
Nodes	587,018	1,300,832
Nodes (target type)	472,432	1,296,675
Edges	15,635,642	29,823,525
Features shape	[472432, 390]	[1296675, 9]
Labeled test samples	94,487	259,335
Edge types	105	23

(1) The Vesta Dataset. Vesta company collected this dataset based on its real-world credit card transactions. The raw data contains 432 features ranging from device type to payment card information and product features. We chose a total of 472,432 instances, of which only 3.5% are fraudulent samples. It is a highly imbalanced dataset with a large number of missing values. The CHET-Graph constructed from the Vesta dataset consists of 587,018 nodes, of which 472,432 are the target type. We extracted 105 different edge types and generated 15,635,642 edges in total.

For such a dataset containing enormous missing values, most existing methods will choose to delete the features whose missing value ratio exceeds a specific limit in the data preprocessing stage, and fill the missing values of some remaining features with the mean or median. It results in the loss of a lot of original information. The

CHET-graph we constructed solves this problem well, making full use of the raw information without adding unexplainable padding.

(2) The Sparkov Dataset. This is a Kaggle public dataset generated using the Sparkov Data Generation tool. It contains 1,296,675 transaction records from 1000 customers with a pool of 800 merchants from 1st January 2019 to 31st December 2020, including 7,780 fraud instances. There are 22 raw features representing information on merchants, customers, and transaction categories. The CHET-graph generated from the Sparkov dataset has 1,300,832 nodes containing 1,296,765 target-type nodes. 29,823,525 edges of 23 different types are included in the final graph.

3.4.2 Compared Methods

We employ several state-of-the-art methods on our selected datasets to verify the effectiveness of our proposed method in credit card fraud detection. These baselines include:

- LR (Itoo et al., 2021): A typical linear model widely used in financial applications.
- KNN (Itoo et al., 2021): A supervised machine learning model that makes predictions based on the similarity between new samples and available samples.
- RF (Kumar et al., 2019): A machine learning classifier that combines the outputs of multiple decision trees for prediction based on bagging and feature randomness methods.
- GBDT (Ge et al., 2020): A machine learning model that ensembles weak predictive models in an iterative optimization process.
- XGBoost (Meng et al., 2020): Extreme Gradient Boosting, an optimized distributed machine learning model under the Gradient Boosting framework.

- CNN (Fu et al., 2016): A convolution neural network model that applies a two-layer convolution.
- LSTM (Jurgovsky et al., 2018): Phrase the fraud detection problem as a sequence classification task and employ LSTM networks to incorporate transaction sequences.
- FIW-GNN: Our proposed method. We construct three variants of the FIW-GNN to comprehensively compare and analyze the best method for assigning edge weights. They are,
- FIW-GNN-n: A FIW-GNN model with no edge weights, which means all the edge weights are equal to 1.
- FIW-GNN-c: A FIW-GNN model with edge weights using the Chi-Squared method. We use the Chi-Squared feature selection method to generate the feature importance values among the selected feature corresponding to different edge types in the CHET-graph.
- FIW-GNN-m: A FIW-GNN model with edge weights using the mutual information method. We use the mutual information feature selection method to generate the feature importance values among the selected feature corresponding to different edge types in the CHET-graph.

3.4.3 Evaluation Metrics

A significant phenomenon of credit card transaction datasets is that they are highly imbalanced. Therefore, in the classification task of detecting fraudulent behaviors, the evaluation metrics we choose should not be biased towards any class (Luque et al., 2019). In this study, we select three evaluation metrics suitable for imbalanced datasets, F1-score, AUC, and G-Mean, to compare the fraud detection performance of all models.

The first metric, F1-score, evaluates a model's binary classification, defined as the harmonic mean of the model's precision and recall. A model's precision and recall can be obtained through the classic confusion matrix, which provides a tabular summary of actual and predictive labels. F1-score can be calculated under (3.5), and the precision and recall can be calculated using (3.6) and (3.7).

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.5}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.6}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.7}$$

The second metric, AUC, is the area under the ROC (receiver operating characteristic) curve. AUC indicates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR), which can be calculated by (3.8) and (3.9).

$$TPR = \frac{TP}{TP + FN} \tag{3.8}$$

$$FPR = \frac{FP}{FP + TN} \tag{3.9}$$

The third metric, G-Mean, is defined in (3.10). It is the geometric mean of sensitivity and specificity, and could be calculated by the square root of the True Positive Rate (TPR) and True Negative Rate (TNR) multiplication. It aims to maximize the accuracy of both classes while keeping them balanced.

$$G\text{-}Mean = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP}} \tag{3.10}$$

All three metrics above are helpful and vital in dealing with the imbalanced dataset. The larger the value of the F1-score, AUC, and G-mean, the better the model's prediction performance.

3.4.4 Implementations and Parameter Settings

The proposed framework FIW-GNN is implemented under PyTorch (v1.12.1), DGL (v0.9.1), and NumPy (v1.21.5). All included datasets are split into training and testing sets with an 80:20 ratio. For all classic machine learning and deep learning models, we remove features with a missing value ratio of more than 20%, while in the graph-based models, we keep all the raw features as they can handle the missing value issues properly. We use the same experimental setup and hyperparameter settings for all graph-based methods. We set the learning rate as 0.01 and the weight decay as 0.0005. All learnable model parameters are optimized with the Adam optimizer. All graph-based models are trained with hidden dimensions of 16, and the number of epochs for training is set to 200 with a dropout rate of 0.2.

3.5 Experimental Results and Discussion

3.5.1 Performance Comparison

To verify that our proposed FIW-GNN can outperform the state-of-the-art methods in the detection performance of credit card fraud, we run all the models on the Vesta and Sparkov datasets, respectively, and calculate the corresponding evaluation metrics for comparison. The corresponding F1-score, AUC, and G-Mean are reported in Table 3.2 and Table 3.3, respectively. We can further divide and analyze these results into three approaches.

As shown in Table 3.2, the proposed FIW-GNN variants consistently outperform traditional machine learning and deep learning baselines on the Vesta dataset, particularly in terms of F1-score and G-Mean.

Similarly, Table 3.3 demonstrates that FIW-GNN achieves the best overall performance on the Sparkov dataset, with FIW-GNN-c yielding the highest scores across all evaluation metrics.

First, in the fraud detection performance of both datasets, the machine learn-

Table 3.2: Performance comparison on the Vesta dataset for fraud detection.

Method	**F1-Score**	**AUC**	**G-Mean**
LR	0.3802	0.8286	0.5029
KNN	0.3656	0.7182	0.4949
RF	0.4568	0.8950	0.5536
GBDT	0.3989	0.7039	0.5219
XGBoost	0.4476	0.8843	0.5524
CNN	0.4053	0.8686	0.5096
LSTM	0.4383	0.8676	0.5478
FIW-GNN-n	0.4767	0.9005	0.5802
FIW-GNN-c	0.5230	0.9005	0.6146
FIW-GNN-m	0.5006	0.9011	0.5955

Table 3.3: Performance comparison on the Sparkov dataset for fraud detection.

Method	**F1-Score**	**AUC**	**G-Mean**
LR	/	0.6725	/
KNN	0.1673	0.6788	0.3161
RF	0.4128	0.8458	0.5218
GBDT	0.3534	0.8759	0.4729
XGBoost	0.4034	0.9210	0.5167
CNN	0.2076	0.8901	0.3408
LSTM	0.4835	0.6864	0.5710
FIW-GNN-n	0.4813	0.9524	0.6026
FIW-GNN-c	0.5139	0.9627	0.6155
FIW-GNN-m	0.4654	0.9480	0.5802

ing models' results are inconsistent. In the Vesta dataset, RF is the best model overall. It ranks first in F1-score, AUC, and G-Mean results. The performance of XGBoost ranks second, which is very close to the results of RF. In the Sparkov dataset, the performance of LR, KNN, and GBDT is not as good as their results in the Vesta dataset, which indicates that these models lack robustness in different datasets. XGBoost and RF perform significantly better than the other machine learning models. XGBoost is similar to RF on F1-score and G-Mean, but its AUC is improved from 0.8458 to 0.9210. Therefore, among all five machine learning models, RF and XGBoost are the models with relatively good overall fraud detection performance.

Second, the deep learning models have achieved more stable fraud detection performance compared with the machine learning models. But there are still fluctuations that cannot be ignored for the two datasets. For example, as shown in Table 3.2, CNN and LSTM perform better than most machine learning models on the Vesta dataset. LSTM is better than CNN on F1-score and G-Mean. However, as shown in Table 3.3, on the Sparkov dataset, CNN performs poorly on other metrics despite achieving good AUC. LSTM surpassed the best RF in machine learning models on F1-score and G-Mean, reaching 0.4835 and 0.5710, respectively, but its AUC is not ideal. We believe that using corresponding feature engineering techniques for different datasets can improve the detection performance of these deep learning models, but this will also lead to a lack of scalability.

Third, compared with the first two approaches of methods, the GNN-based method has been significantly improved in F1-score, AUC, and G-Mean on both datasets. For example, in the results of the Vesta dataset, the FIW-GNN-c method improved by 14.49% and by 11.02% on G-Mean compared to the previous best RF. The performance of FIW-GNN-m on the Vesta dataset is also outstanding, achieving an AUC of 0.9011, F1-score of 0.5006, and G-Mean of 0.5955, which is slightly better than FIW-GNN-c in AUC, but inferior in F1-score and G-Mean. In the results of the Sparkov dataset, we can also draw the same conclusion. All three GNN-based

models achieve better AUC and G-Mean than the other methods, and FIW-GNN-c achieves the highest F1-score with a result of 0.5139. The results show that the GNN-based methods have good robustness on both datasets. In the histograms of evaluation metrics for both datasets, the GNN-based methods outperform the best-compared models in the other two approaches on F1-score, AUC, and G-Mean, as shown in Fig. 3.4 and Fig. 3.5.

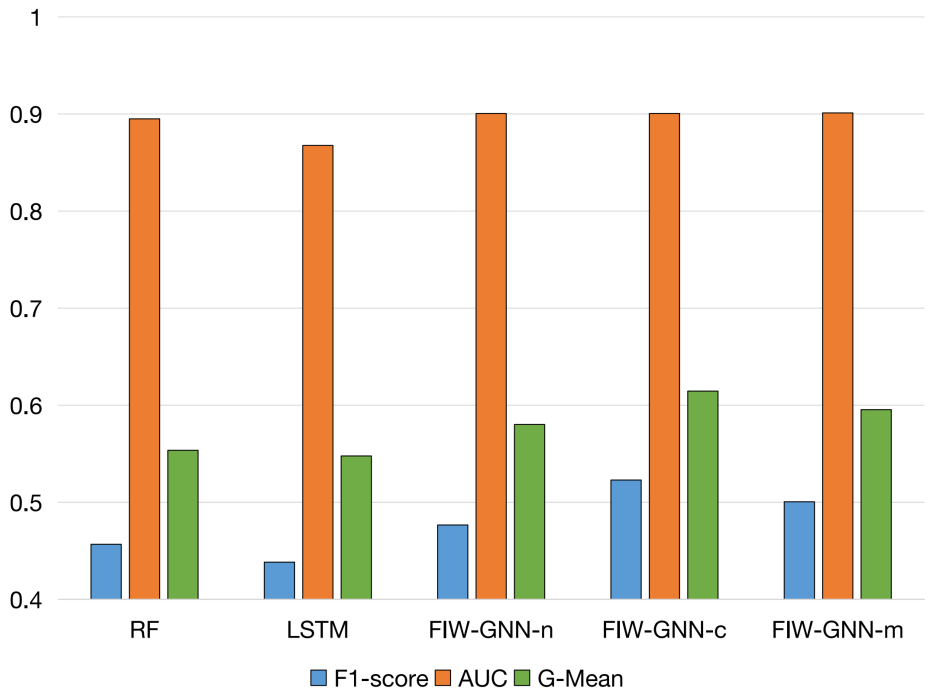


Figure 3.4: Histogram of the model evaluation metrics on the Vesta dataset.

As a complementary comparison, Fig. 3.5 presents the corresponding results on the Sparkov dataset.

In summary, the GNN-based methods outperform traditional machine learning and deep learning methods in solving the problem of credit card fraud detection. Moreover, our proposed FIW-GNN-c method is the overall winner in F1-score, AUC, and G-Mean among all ten compared methods under both transactional datasets.

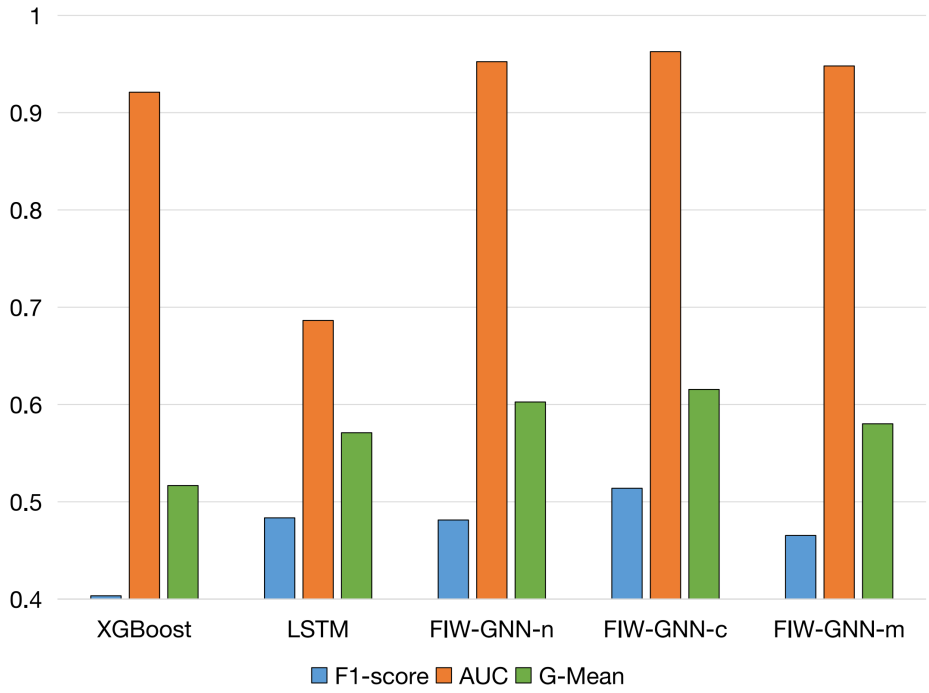


Figure 3.5: Histogram of the model evaluation metrics on the Sparkov dataset.

3.5.2 Ablation Study

We verify the effectiveness of our model components from two perspectives through an ablation study.

The first perspective is to verify the effectiveness of the feature importance-based edge weights by removing them, and to compare the effects between the Chi-squared statistic method and the mutual information statistic method. As shown in Table 3.2, the models using edge weights based on feature importance is significantly better than the FIW-GNN-n model without edge weights. In particular, the FIW-GNN-c model using the Chi-square method improves F1-score and G-Mean by 9.71% and 5.93% compared to FIW-GNN-n. Similarly, the FIW-GNN-m model using the mutual information method also outperforms FIW-GNN-n on the three evaluation metrics. The results of the Sparkov dataset in Table 3.3 show that the performance of using the mutual information method is slightly lower than that of not using edge weights. However, the FIW-GNN-c model using the Chi-square method shows good robustness and is also better than FIW-GNN-n in three evaluation metrics.

In summary, we can conclude that using feature importance-based edge weights can effectively improve the detection performance of the model. Besides, the FIW-GNN-c model using the Chi-square method is robust and significantly enhances the performance of both datasets. Therefore, our final model adopts edge weights assignment based on the Chi-square statistic method.

The second perspective concerns the effects of different edge weight ranges on the three evaluation metrics. After obtaining the values of feature importance through Chi-squared or Mutual information method, we use MinMaxScaler to scale them to a specified range. Table 3.4 shows the impact of different edge weight ranges on the evaluation metrics. It shows that as the value range increases, the performance of the evaluation metrics will not be improved synchronously. Especially for the Vesta dataset with higher dimensions, whether it is the Chi-square or mutual information method, the optimal results are achieved on a minor value interval (1, 2). Therefore, we use (1, 2) as the range of edge weights in the final FIW-GNN model.

3.6 Conclusions and Future Work

In this chapter, we propose a feature importance-based weighted heterogeneous graph neural network named FIW-GNN to solve the problem of credit card fraud detection. In our FIW-GNN model, we propose a method to construct a novel heterogeneous graph for credit card transactional datasets, namely CHET-graph. The CHET-graph is suitable for real-world credit card transactional data, and it can competently deal with the problems of the existence of enormous missing values, lack of objective definition standards, and poor adaptability among different credit card transaction datasets. Moreover, based on the architecture of RGCN, we propose a feature importance-based method to assign edge weights. Finally, we evaluate our proposed method on two benchmark credit card transaction datasets compared to other state-of-the-art models. The results show that FIW-GNN achieves the best

Table 3.4: Evaluation metrics of different edge weight ranges in the ablation study.

Dataset	Method	Range	F1-Score	AUC	G-Mean
Vesta	**Chi-squared**	(1,2)	0.5230	0.9005	0.6146
		(1,3)	0.4761	0.8984	0.5778
		(1,5)	0.4739	0.8987	0.5691
	Mutual information	(1,2)	0.5006	0.9011	0.5955
		(1,3)	0.4760	0.9010	0.5707
		(1,5)	0.4054	0.8906	0.5124
Sparkov	**Chi-squared**	(1,2)	0.5139	0.9627	0.6155
		(1,3)	0.5576	0.8991	0.6469
		(1,5)	0.5434	0.9687	0.6408
	Mutual information	(1,2)	0.4654	0.9480	0.5802
		(1,3)	0.4716	0.9130	0.5835
		(1,5)	0.5320	0.9149	0.6281

detection performance on both datasets compared to other competing approaches, demonstrating our proposed method's effectiveness.

For future work, exploring different combinations, patterns, and relations between feature importance and heterogeneous graph construction will help to further increase the interpretability of the graph and further improve the performance in credit card fraud detection. Meanwhile, based on GNN methods, solving other potential problems in transactional datasets, such as fraudulent camouflage and fairness test (Y. Dong et al., 2022), can also be a promising direction for credit card fraud detection.

Despite the promising performance achieved on the benchmark datasets, several practical limitations should be acknowledged when considering real-world deployment of the proposed framework. First, the study is conducted under a transductive learning setting using a static transaction graph, whereas real-world fraud detection systems often require continual adaptation to evolving transaction patterns. Sec-

ond, although the graph construction strategy is designed to capture meaningful transactional relationships, its effectiveness may depend on the characteristics of the underlying financial network and therefore warrants further evaluation across diverse datasets. Addressing these limitations through inductive graph learning, temporal graph modeling, and evaluation under streaming environments would further improve the scalability and practical applicability of the proposed framework.

Chapter 4

Balancing Group and Individual Fairness in Graph Neural Networks

Graph neural networks (GNNs) nowadays play a vital role in many high-stakes applications due to their powerful data representation capabilities. In recent years, the issue of fairness discrimination in GNNs has gained significant attention. Existing methods endow different concepts of fairness in GNNs, mainly focusing on group and individual fairness, and give corresponding solution frameworks. However, these existing methods usually only consider optimizing one kind of fairness and claim that group and individual fairness are conflicting and cannot be handled together. Such an approach has limitations and lacks practicality for real-world application scenarios. In this chapter, we propose a GNN-based framework for resolving the conflict between group and individual fairness, namely **BIG** (**B**alancing **I**ndividual and **G**roup fairness). First, we expound on the view that group and individual fairness do not conflict and define a novel notion named balanced fairness and the corresponding evaluation metric. Second, we propose the framework structure and algorithm principle of BIG. Finally, we evaluate the effectiveness of BIG on three real-world datasets. The experimental results show that BIG balances group fairness and individual fairness in GNNs well while ensuring the accuracy of the node classification tasks.

4.1 Introduction

In recent years, graphs, as a nonlinear data structure that can effectively represent data and the relationship between instances, have been used to solve many real-life problems. For example, social networks (Fan et al., 2019), financial risk control (Sarkar & Alvari, 2020), knowledge graphs (Ji et al., 2021), etc. In machine learning, based on the powerful expressiveness of the graph structure, graph neural networks (GNNs) that specialize in processing graph domain information have been rapidly developed and widely used because of their superior performance and interpretability. GNNs have made remarkable breakthroughs in recommender systems (Ying et al., 2018), natural language processing (Yao et al., 2019), intelligent cities (G. Dong et al., 2023), and more. Although GNNs have shown excellent predictive performance in the above applications, the issue of fairness discrimination in the use of GNNs has recently attracted the academic community's attention. This has led to discussions about the potential trade-offs between different types of fairness (Y. Dong et al., 2023)(Dai & Wang, 2021).

In order to achieve fairness in machine learning, many research works have introduced fairness measures in the evaluation of machine learning models to solve the bias in the training data and further correct corresponding problems caused by bias or discrimination (Mehrabi et al., 2021)(Pessach & Shmueli, 2022). This improvement in the performance of machine learning models stops them from making wildly biased predictions for certain groups of people, which can improve overall fairness. For traditional machine learning models, some studies modify the labels and attributes of the training dataset in the data preprocessing stage to eliminate bias and obtain a non-discriminatory fair-aware dataset (Le Quy et al., 2022). There are also some studies to make the model generate fair prediction results by adding regular term or loss functions based on fairness to the machine learning model (Agarwal et al., 2021). However, with the rise and widespread use of GNNs, research on their fairness is still in its infancy. Many issues must be resolved, such as the definitions

of fairness concepts, fairness performance measures, and fairness conflict.

In general, in the fairness of machine learning, there are two main types of mainstream concepts: group fairness and individual fairness. Group fairness emphasizes equal treatment across demographic subgroups (Hardt et al., 2016), often perceived as conflicting with individual fairness, which focuses on similar treatment for similar individuals (Dwork et al., 2012). The research on fairness in GNNs mainly revolves around the above two concepts. There are many discussions on group fairness in GNNs (Dai & Wang, 2021) (Agarwal et al., 2021), while research on individual fairness in GNNs is relatively insufficient. In addition, these two types of research often claim each other's limitations in considering fairness issues, which makes the two types of views always exist in a conflicting manner (Binns, 2020).

To address the abovementioned issues, we investigate a novel problem in this chapter, how to balance the conflict between group and individual fairness that arises in GNNs. First, we present a novel perspective that challenges the traditional view of an inherent conflict between group and individual fairness. We argue that these concepts can be balanced effectively, allowing for a more comprehensive approach to fairness. In addition, considering only one kind of fairness and completely ignoring the other kind of fairness is unsuitable for real-world application scenarios. Then, we propose a GNN-based framework, BIG (Balancing Individual and Group fairness), to resolve the conflict between group and individual fairness. Moreover, we defined a new concept of balanced fairness and designed a corresponding evaluation metric, balanced fairness score. To verify the effectiveness of the BIG framework on fairness problems, we conduct experiments on two real-world datasets and compare them with the existing state-of-the-art fairness methods in GNNs.

The main contributions of our research can be summarized as follows:

- We introduce a novel perspective on the fairness problem in GNNs, showing that group and individual fairness can be harmonized based on the application context, which is crucial for real-world scenarios. It is essential to balance both

aspects without neglecting either.

- We design a novel GNN-based fairness framework, BIG, to balance group fairness and individual fairness issues in GNNs.
- We define a new concept of fairness, balanced fairness, and design a corresponding evaluation metric, balanced fairness score.
- We conduct comparative experiments based on three real-world datasets and verify our proposed framework's effectiveness and superior performance based on our selected and proposed evaluation metrics.

The remainder of this chapter is organized as follows. Section 4.2 presents a literature review on the development of GNNs in recent years and the research status of group fairness and individual fairness in GNNs. Section 4.3 conducts the preliminaries to understand the fairness problem in GNNs. Section 4.4 proposes our novel GNN framework BIG to solve the fairness issue. Section 4.5 introduces the experiments on three real-world datasets and discusses the experimental results, and Section 4.6 summarizes the chapter's conclusions and future work.

4.2 Related Work

Research on fairness in machine learning has evolved over the years. GNN, as one of the most popular branches in the field of deep learning recently, research on its fairness issues is still at the beginning stage. We summarize the related work on the fairness problem in GNNs through three approaches: 1) Graph neural networks approach, 2) Group fairness in GNNs approach, and 3) Individual fairness in GNNs approach.

4.2.1 Graph Neural Networks

In recent years, Graph neural networks (GNNs) have become an area of rapid development in deep learning. It has achieved significant breakthroughs in various large-scale application models, for example, recommender systems (L. Wu et al., 2022), traffic forecasting (Jiang & Luo, 2022), drug discovery (Xiong et al., 2021), and protein discovery (Strokach et al., 2020).

The topological structure formed by the nodes in the graph and the characteristic information of the nodes is represented as an embedding space, and the GNN uses pairwise message passing (Hamilton et al., 2017) to exchange information between nodes and their neighbors to update the embedding representation continuously iteratively. Kipf et al. (Welling & Kipf, 2017) extended convolution operations to graph data and proposed a graph convolutional network (GCN), a popular GNN. The core is to learn a function mapping so that the nodes in the graph can aggregate their characteristics and neighbor characteristics and then generate a new representation of the node. Based on GCN, a relational graph convolutional network (R-GCN) that specializes in dealing with highly multi-relational data features has also evolved (Schlichtkrull et al., 2018). On the other hand, GraphSAGE (Hamilton et al., 2017) optimizes the traditional GCN model, develops a general inductive framework, and endows the graph model with the ability of inductive learning. It optimizes the sampling of the entire graph to the sampling of the current neighbor node, continuously aggregates neighbor information, and then iteratively updates. Veličković et al. (Velickovic et al., 2017) used attention in the graph data learning process to amplify the influence of essential parts of the data. They proposed a graph attention network (GAT), which can adaptively learn the importance weights of neighbor nodes. Moreover, Xu et al. (Xu et al., 2018) further expanded the discrimination of GNNs on different simple graph structures by adaptively adjusting the importance weights of different nodes and proposed Graph Isomorphism Network (GIN).

However, the GNNs transfer information in the graph structure to obtain representations also involves sharing and transferring sensitive features. It will raise the problem of fairness discrimination in GNNs. In particular, because graph structures can represent and convey richer information than traditional datasets, which can further exacerbate fairness discrimination. Existing research on GNNs has achieved considerable improvements in the performance of practical applications but neglects to identify sensitive information in real-world data and deal with fairness issues (Y. Dong et al., 2023). Research on solving the fairness problem in GNNs has recently appeared (Agarwal et al., 2021) (Dai & Wang, 2021), but it is still exploratory. There is a lack of uniform standards for the definition of fairness in GNNs, and there are widespread disputes about the tendency of different types of fairness (Burkholder et al., 2021).

4.2.2 Group Fairness in GNNs

Group fairness means that for subgroups divided by different attributes of sensitive features, the prediction results generated by the algorithm should not produce discriminatory bias (Berk et al., 2021). Group fairness mainly appears in high-risk applications, such as recidivism prediction (Dressel & Farid, 2018), loan evaluation (D. Wang et al., 2019), etc. Sensitive information, such as race, skin color, and gender, is protected by law and should not be used as the basis for decision-making. There are three popular notions of group fairness: demographic parity, equality of odds, and equality of opportunity.

Based on demographic parity, Dai et al. (Dai & Wang, 2021) proposed FairGNN to complete fair node classification for binary-sensitive features in the dataset. In addition, FairGNN also considers the problem of a large number of missing sensitive information encountered in real-world datasets. Similarly, Agarwal et al. (Agarwal et al., 2021) designed a GNN framework called NIFTY that considers group fairness and stability. On the one hand, based on counterfactual fairness, it generates another

graph structure in which sensitive information of all nodes is flipped, so its global prediction result is close to the original graph. On the other hand, it enhances the stability of NIFTY by adding disturbing noise to the three dimensions of node features, sensitive features, and graph structure. Equality of odds means that in the binary classification task, under the condition of ground truth labels, the prediction results should be independent of sensitive features (Hardt et al., 2016). Ma et al. (J. Ma et al., 2021) developed a PAC-Bayesian analysis in GNN under semi-supervised learning, which makes its GNN have better generalization performance in the equality of odds in node classification. Equality of opportunity is an extension of the equality of odds. It means that under ground truth labels, the positive prediction results should be independent of sensitive features (Hardt et al., 2016). A Bayesian approach called DeBayes was proposed by Buyl et al. (Buyl & De Bie, 2020) to learn debiased embeddings using a biased prior, which can simultaneously achieve excellent demographic parity and equalized opportunity in link prediction tasks in graph structures.

Incorporating group fairness into the framework of GNNs can effectively improve the practicability of GNNs in the real world. However, group fairness looks at sensitive characteristics and eliminates discrimination from a relatively macro perspective, and it will not be able to deal with finer-grained fairness issues (Kang et al., 2020). It has led to limitations in application scenarios where the sample space is not large enough or a specific need for fairness in a finer dimension.

4.2.3 Individual Fairness in GNNs

Individual fairness is at the individual level compared to group fairness and does not consider a specific sensitive characteristic. Roughly speaking, individual fairness was initially defined by Dwork et al. as, in classification, similar individuals should be treated similarly (Dwork et al., 2012). So far, individual fairness has been applied in several studies on GNNs.

Kang et al. (Kang et al., 2020) debiased the input graph, graph mining model, and graph mining results through three complementary algorithm frameworks, thus constructing an individual fairness method called InFoRM for multi-graph mining tasks. Lahoti et al. (Lahoti et al., 2019) think that the original definition of individual fairness is difficult to measure by similarity metric, so they redefine individual fairness through a unified Pairwise Fair Representation (PFR). It captures the data-driven similarity between individuals and pairwise edge information in fairness graphs. Similarly, a ranking-based framework, REDRESS, is proposed to redefine individual fairness from a ranking perspective (Y. Dong et al., 2021). It encapsulates the predictive performance of the GNN model and the ranking-based individual fairness as an optimization problem to achieve a good balance. In addition, Wang et al. (Y. Wang et al., 2022) introduced individual fairness to the application of multi-view graphs, which contain multiple single-view graphs with the same node set but different edge types. Dong et al. (W. Song et al., 2022) considered that there are limitations in measuring individual fairness through an overall evaluation method and proposed the framework of GUIDE. GUIDE not only considers the overall individual fairness but also considers the individual fairness of subgroups according to sensitive attributes and optimizes the individual fairness of different subgroups to be as consistent as possible.

However, compared with the studies considering group fairness in GNNs, the number of studies considering individual fairness in GNNs is still minimal. A significant problem here is that individual fairness is more complicated to define than group fairness because it does not depend on any specific sensitive feature. Several studies mentioned above (Lahoti et al., 2019)(Y. Dong et al., 2021) have proposed their definitions for individual fairness. On the other hand, many studies believe there is a conflict between group and individual fairness, and it is impossible to make an effective trade-off (Binns, 2020). Therefore, more research addresses group fairness with a more straightforward definition. Nevertheless, in real-world application scenarios, individual fairness has important implications in practical applications

and cannot be ignored.

4.3 Preliminaries

In this section, we first introduce the notations and preliminaries of GNNs. Then, we illustrate the concept of Balanced Fairness and list the methods to measure it. Finally, we formally give the problem definition.

4.3.1 Notations

In this chapter, an attributed graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, which consists of: (1) $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$, the set of nodes ($|\mathcal{V}| = n$), where v_i is the i-th node in the graph $\mathcal{G}$, (2) $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, the set of edges, (3) $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, and $\mathrm{X} \in \mathbb{R}^{N \times d}$, the set of node features, where $\mathbf{x}_i \in \mathrm{R}^d$ is the attribute vector of the i-th node.

$\mathbf{A} \in \{0, 1\}^{N \times N}$ is the adjacency matrix of the graph $\mathcal{G}$, where $\mathbf{A}_{ij} = 1$ if nodes v_i and v_j are connected; otherwise, $\mathbf{A}_{ij} = 0$. We assume there is a sensitive attribute $\mathrm{s} \in \{0, 1\}$. The set of sensitive attribute is denoted as $\mathcal{S}$. $\mathcal{S}$ yields two disjoint groups in $\mathcal{V}$. We use $\mathcal{V}_{(s=0)}$ to denote the set of nodes in the group whose sensitive attribute's value equals to 0. We also assume there is a similarity matrix $\mathbf{W}$, which describes pairwise similarities of nodes in the input space. The greater similarity of two nodes, the greater the value of the measure. $y_v \in \mathcal{Y}$ is the ground truth label of node v, and $\hat{y_v} \in \hat{\mathcal{Y}}$ is the predicted label of node v.

4.3.2 Preliminaries of Graph Neural Networks

In this chapter, GNNs aim to predict the label y_v of each node v. GNNs utilize node attributes and edges to learn a representation embedding $\mathbf{h}_v$ for node v. For each node embedding $\mathbf{h}_v$, apply a linear classifier to make binary predictions on nodes as $y_v = f(\mathbf{h}_v)$. Current GNNs use message passing, where adjacent nodes or edges exchange information and influence each other's updated embeddings. For each node

v in the graph, collect the embeddings or messages of all neighboring nodes. Then, aggregate the messages of all neighboring nodes through the aggregation function. All pooled messages are passed through the update function, which is iteratively updated for k layers. The update process of the k-th layer in GNN can be expressed as follows:

$$\mathbf{a}_v^{(k)} = \text{AGGREGATE}^{(k-1)}\left(\left\{\mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v)\right\}\right), \tag{4.1}$$

$$\mathbf{h}_v^{(k)} = \text{COMBINE}^{(k)}\left(\mathbf{h}_v^{(k-1)}, \mathbf{a}^{(k)}\right), \tag{4.2}$$

where $\mathcal{N}(v)$ is the set of neighboring nodes of v, and $\mathbf{a}_v^{(k)}$ is the representation of node v's neighboring nodes at the k-th layer.

4.3.3 Balanced Fairness

In this subsection, we introduce a novel fairness concept proposed in this chapter, **Balanced Fairness**.

Many existing studies on fairness in GNNs are limited to solving one kind of fairness problem: group or individual fairness. Moreover, the relationship between these two kinds of fairness is often described as conflicting (Y. Dong et al., 2023). We argue that the conflict between group and individual fairness is based on a misunderstanding. Apparently, they are not mutually exclusive regarding definition or plausible motivation. Moreover, from the perspective of political and legal philosophy (Binns, 2020), individual fairness and group fairness do not conflict fundamentally.

Therefore, (1) group and individual fairness are not mutually conflicting or exclusive. They consider fairness in machine learning at two different granularities. (2) For the fairness problem in GNNs, we need to consider both group and individual fairness. Only considering one kind of fairness is not in line with real application scenarios and needs in the real world. (3) The balance between group fairness and individual fairness and the determination of the proportion between the two should

be based on business goals or application scenarios rather than on the difficulty of achieving fairness, the priority of model utility, or the fairness measures.

Based on the above viewpoints, we define a novel concept of fairness, Balanced Fairness, which explains the goal of meeting both group and individual fairness and better combines real-world application scenarios and business needs. We can formally define Balanced Fairness as:

Definition 4.3.1. Balanced Fairness is satisfied if the algorithm does not yield discriminatory predictions against (1) individuals from any specific sensitive subgroups and (2) any pair of individuals with sufficiently high similarity.

4.3.4 Measuring Balanced Fairness

In this subsection, we introduce the metrics that can be used to measure the level of group and individual fairness in GNNs. Among them, for individual fairness, we define another novel metric, **Individual Fairness Parity (IFP)**, to ensure the robustness of individual fairness in GNNs during the measurement process. Subsequently, we formally define a metric, **Balanced Fairness Score (BFS)**, which can be used to quantify the Balanced Fairness in GNNs. Here, we only consider binary node classification tasks in GNNs. We denote $y \in \{0, 1\}$ as the binary label, $s \in \{0, 1\}$ as the sensitive attribute, and $\hat{y} \in \{0, 1\}$ as the prediction.

4.3.4.1 Measuring Group Fairness

In node classification, the assessment of group fairness usually uses a parity measure, which is a static observational criterion. It requires the evaluation metric to be independent of the salient groups. In BIG, we use two commonly used criteria to measure group fairness: statistical parity (Dwork et al., 2012) and equal opportunity (Hardt et al., 2016). These criteria are reasonable for uncovering potential inequities and are particularly useful in monitoring real-time decision-making systems. They

can be defined as:

Definition 4.3.2. Statistical Parity is satisfied if the predictions are independent for individuals in both sensitive subgroups. It can be formally formulated as:

$$P(\hat{y} \mid s = 0) = P(\hat{y} \mid s = 1), \tag{4.3}$$

Definition 4.3.3. Equal Opportunity is satisfied if the positive predictions are independent for individuals in both sensitive subgroups. It can be formally formulated as:

$$P(\hat{y} = 1 \mid y = 1, s = 0) = P(\hat{y} = 1 \mid y = 1, s = 1), \tag{4.4}$$

To quantitatively evaluate statistical parity and equal opportunity in our proposed framework, we utilize the following metrics, Δ_{SP} and Δ_{EO}, to measure how far the prediction deviates from the ideal situation that satisfies statistical parity and equal opportunity:

$$\Delta_{SP} = |P(\hat{y} = 1 \mid s = 0) - P(\hat{y} = 1 \mid s = 1)|, \tag{4.5}$$

$$\Delta_{EO} = |P(\hat{y} = 1 \mid y = 1, s = 0) - P(\hat{y} = 1 \mid y = 1, s = 1)|, \tag{4.6}$$

Generally, smaller values of Δ_{SP} and Δ_{EO} indicate that a higher level of group fairness is achieved.

4.3.4.2 Measuring Robust Individual Fairness

Compared with group fairness, individual fairness is at a finer-grained level, which does not consider any sensitive features but only focuses on each node in the graph. First, we use a pairwise node distance method based on Lipschitz Condition (Kang et al., 2020)(Lahoti et al., 2019) to evaluate the overall individual fairness. In graph mining, Lipschitz Condition is defined as:

Definition 4.3.4. Lipschitz Condition is satisfied if the distance between any pair of nodes in the output space is less than or equal to their distance in the input space. It can be formally formulated as:

$$\mathbf{D}\left(f\left(v_i\right), f\left(v_j\right)\right) \leq \mathbf{L} \cdot \mathbf{d}\left(v_i, v_j\right), \tag{4.7}$$

where v_i and v_j are two nodes in the graph, $f(\bullet)$ is a function that gives the output node embeddings, $\mathbf{D}(\bullet)$ and $\mathbf{d}(\bullet)$ are the distance metrics in the output and input space respectively, and $\mathbf{L}$ is called the Lipschitz constant.

We utilize a metric called *Consistency* (Lahoti et al., 2019) to evaluate the overall individual fairness based on Lipschitz Condition quantitatively. It can be formulated as follows based on the pairwise similarity matrix $\mathbf{W}$:

$$Consistency = 1 - \frac{\sum_{v_i} \sum_{v_j} |\hat{y}_i - \hat{y}_j| \cdot \mathbf{W}[i,j]}{\sum_{v_i} \sum_{v_j} \mathbf{W}[i,j]} \quad \forall i \neq j, \tag{4.8}$$

where $\mathbf{W}[i,j]$ represents the (i,j)-th entry of $\mathbf{W}$. Generally, a larger value of *Consistency* indicates that a higher level of overall individual fairness is achieved.

To keep synchronicity with other fairness metrics in this framework, which is that the smaller the value, the better it is to achieve the fairness goal, we define *Inconsistency*, which equals $1 - Consistency$.

The overall individual fairness is to judge fairness from a finer-grained level. However, evaluating from a finer-grained level will lead to a lack of robustness. Here, to improve the robustness of individual fairness evaluation, in addition to quantitative statistics from the whole, we can also perform quantitative statistics from different subgroups. The original intention here is that if individual fairness is good enough, the individual fairness metrics between different subgroups should also be the same. Therefore, we define Individual Fairness Parity as:

Definition 4.3.5. Individual Fairness Parity (IFP) is satisfied if the consistencies of different sensitive subgroups are equal. It can be formally formulated

as:

$$Consistency_{(s=0)} = Consistency_{(s=1)}, \tag{4.9}$$

To quantitatively evaluate the Individual Fairness Parity, we utilize Δ_{IFP}, to measure the difference of consistency between the sensitive subgroups:

$$\Delta_{IFP} = |Consistency_{(s=0)} - Consistency_{(s=1)}|, \tag{4.10}$$

Generally, a smaller value of Δ_{IFP} indicates that a higher level of individual fairness parity is achieved.

Therefore, we construct a more robust individual fairness evaluation system based on the evaluation from the two perspectives of overall individual fairness and individual fairness parity. In this chapter, we call it **Robust Individual Fairness**.

4.3.4.3 Balanced Fairness Score

According to the definition of balanced fairness, the group and individual fairness evaluation method in GNNs, we present an evaluation metric to measure balanced fairness, namely the **Balanced Fairness Score (BFS)**. It is a comprehensive embodiment of group and robust individual fairness measurements, which can be formulated as:

$$\mathbf{BFS} = \Delta_{SP} + \Delta_{EO} + \gamma \cdot (Inconsistency + \Delta_{IFP}), \tag{4.11}$$

where γ controls the exponential difference in the value scale between group and individual fairness metrics that may appear in different datasets and specific experiments.

4.3.5 Problem Definition

Based on our preliminary analysis, with the notations given above, we formally define our research problem as:

Problem 1. Balancing Individual and Group Fairness in GNNs. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{X})$, with corresponding ground truth labels $\mathcal{Y}$, sensitive attributes $\mathcal{S}$, and similarity matrix $\mathcal{W}$ for nodes in $\mathcal{V}$, our goal is to learn a GNN for binary node classification satisfying: (1) Individual fairness and group fairness are considered at the same time; (2) The levels of individual fairness and group fairness are optimized and balanced; (3) The model maintains high accuracy whilst satisfying the fairness criteria.

4.4 Proposed Framework

In this section, we present the details of a novel GNN-based framework for fairness issues, namely **BIG** (**B**alancing **I**ndividual and **G**roup fairness). We first present the overview of BIG. Then, we illustrate the workflow of BIG. Finally, we give the formulation of BIG's objective functions.

4.4.1 Framework Overview

The overview of our proposed framework BIG is shown in Fig. 4.1. First, based on the input graph, a GNN classifier initializes its node embeddings. GNN will complete the classification task of nodes. Second, for each node, we obtain its similarity matrix and the personalized weights generated based on the corresponding neighbor nodes on the similarity matrix. Here, the similarity matrix can reflect that different neighbors of a node have different similarities and group affiliations. In this way, different neighbors can influence the node differently in the final fairness evaluation metrics. Personalized weights are generated using an attention mechanism, exploiting node features and pairwise similarity values. Next, we can get the predicted labels of all nodes generated by the GNN classifier and combine the GNN-initialized node embeddings with the personalized learned weights to get the final output. Finally, to achieve the fairness of the node classification task, we compare the output results with the basic facts and the similarity matrix and realize the

global optimization of node classification prediction, group fairness, and individual fairness according to the loss function defined by the proposed framework.

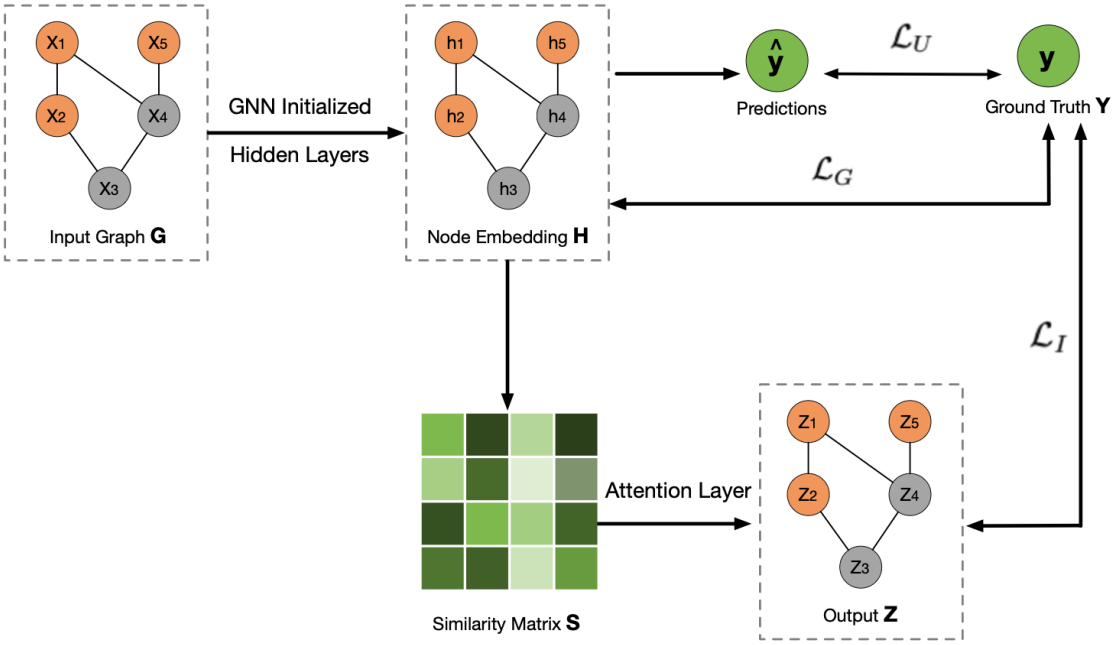


Figure 4.1: The overall framework of BIG.

4.4.2 Workflow of the Proposed Framework

The proposed framework ensures the trade-off between the GNN model's utility and the results' fairness and considers the two sub-goals of the group and individual fairness. Therefore, the workflow of BIG will be carried out according to three main steps, which are (1) to ensure the utility of the GNN model, (2) to enforce group fairness, and (3) to enforce individual fairness.

4.4.2.1 Ensuring the Utility of the GNN Model

We use a backbone GNN model for downstream node classification tasks. In BIG, the main task of the GNN model is to learn a representation vector for each node and complete the prediction of each node's label. First, we extract the graph adjacency matrix $\mathbf{A}$ and node feature matrix $\mathbf{X}$ from the input graph generated from the raw

dataset. Second, the GNN model obtains an initialized node embedding **H**' based on these two matrices. This initialized node embedding **H**' will be processed later for achieving group and individual fairness. Then, the GNN model will iterate through k layers to complete aggregating and combining information between nodes and neighbors. Finally, through a linear classification layer, the GNN model gets the prediction results of the nodes. In order to maintain the utility performance of the GNN model, we use the binary cross-entropy loss function as the objective based on the ground truth y and predicted labels $\hat{y}$ of the nodes.

4.4.2.2 Enforcing Group Fairness

Through the previous step, we obtained the predicted labels of all nodes. In order to achieve the goal of group fairness, based on the group membership of each node and corresponding ground truth label, we conduct quantitative statistics from the perspectives of empirical demographic parity and equal opportunity.

From the perspective of empirical demographic parity, for each type of predicted label, we count the number which belongs to each group. Here, we only consider binary sensitive attributes. For example, for the sensitive attribute $s = 0, 1$, all nodes belong to the group($s = 0$) or group($s = 1$). For the predicted label k, we calculate how many nodes in the two groups have the predicted label k. From the perspective of equal opportunity, for the predicted label k, we already know how many nodes in the two groups have the predicted label k. Then, we calculate the number of nodes predicted to be correct in the number of nodes whose predicted label is k in the two groups. Finally, we optimize through a loss function for group fairness based on these statistics.

4.4.2.3 Enforcing Individual Fairness

The definition of individual fairness is currently controversial, and it is relatively widely accepted that similar individuals should get similar outputs. For GNNs, the prediction results of nodes with similar representations should be similar. Like

other methods designed for individual fairness, BIG aims to maximize the overall individual fairness. We use the GNN model based on the initialized node embedding **H**' obtained in the previous step, and the similarity matrix **W** of the node, and complete the message passing through the aggregation operation. Thus, we obtain the similarity information of the paired nodes.

Here, the difference between enforcing individual and group fairness is that, in the stage of enforcing group fairness, we use the graph adjacency matrix **A** and the node feature matrix **X** as the input for GNN, while in the stage of enforcing individual fairness, we use the node similarity matrix **W** as the adjacency matrix and the node embeddings **H** as the feature matrix.

In addition, in BIG, in order to enhance the robustness of the individual fairness goal, we should also enhance the individual fairness parity between different groups while optimizing the overall individual fairness. We build an attention layer before performing the aggregation operation of the GNN model. Through the attention mechanism of the GNN model, personalized attention weights are established for each node and its neighbors. Attention weights will be learned by optimizing the overall objective function. Finally, we obtain the output embeddings Z of the nodes through the aggregation operation.

4.4.3 Objective Function Formulation

Our proposed framework aims to achieve three optimization goals: (1). maintain the accuracy of the GNN model in node classification prediction, (2). ensure the optimization of overall group fairness, and (3). ensure the optimum of robust individual fairness. Each optimization objective corresponds to a loss function. In addition to this, we also strive to achieve a balance between these three optimization goals. Therefore, we summarize the overall optimization formulation based on these three optimization objectives.

First, for the node classification prediction task of the GNN model, we use the

famous binary cross-entropy loss function as the first optimization item. The binary cross-entropy loss function is widely adopted in deep learning applications and plays an essential role in GNNs. Specifically, it is given as follows:

$$\mathcal{L}_U = \frac{1}{n}\sum_{i=1}^{i=n}\left[-y_{v_i}\log\hat{y}_{v_i} - (1-y_{v_i})\log\left(1-\hat{y}_{v_i}\right)\right], \tag{4.12}$$

where n indicates the size of nodes set $\mathcal{V}$, $y_v^{(i)}$ represents the ground truth label of the i-th node, and $\hat{y}_{v_i}$ represents the predictive label of the i-th node.

Second, BIG uses $\mathcal{L}_G$ as the second optimization term to achieve overall group fairness. As mentioned in Section 4.3, statistical parity and equal opportunity are the two most common metrics to measure overall group fairness in graph learning. Thus, we define $\mathcal{L}_G$ as follows:

$$\mathcal{L}_G = \sum_{j=1}^{c}\left(\frac{\sum_{v_i\in\mathcal{V}_0} P\left(\hat{y}_{v_i}=j\right)}{|\mathcal{V}_0|} - \frac{\sum_{v_i\in\mathcal{V}_1} P\left(\hat{y}_{v_i}=j\right)}{|\mathcal{V}_1|}\right)^2, \tag{4.13}$$

where v_i is the i-th node in the graph; $\mathcal{V}_0$ and $\mathcal{V}_1$ are the node sets for the two sensitive subgroups ($s=0$ and $s=1$), respectively; j is the class label; $\hat{y}_{v_i}$ is the classification result of the i-th node made by the model.

Third, BIG uses $\mathcal{L}_I$ as the third optimization term to achieve robust individual fairness. $\mathcal{L}_I$ consists of two parts: $\mathcal{L}_{oif}$, which guarantees overall individual fairness, and $\mathcal{L}_{ifp}$, which ensures overall individual fairness parity among different groups. Here, the subscript oif means overall individual fairness, and the subscript ifp represents individual fairness parity.

As mentioned in Section 4.3, based on the Lipschitz condition, we use a bias measurement to evaluate the overall individual fairness in graph mining, that is, the bias between the mining results and the similarity matrix. The function of $\mathcal{L}_{oif}$ can be formulated as follows:

$$\mathcal{L}_{oif} = \frac{\sum_{v_i\in\mathcal{V}}\sum_{v_j\in\mathcal{V}}\|\mathbf{Z}[i,:]-\mathbf{Z}[j,:]\|_2^2\mathbf{W}[i,j]}{2}, \tag{4.14}$$

where $\mathbf{Z}$ is the output node embeddings matrix. $\mathbf{Z}[i,:]$ and $\mathbf{Z}[j,:]$ represent the i-th and j-th row of $\mathbf{Z}$, respectively.

In order to maintain more robust individual fairness in our framework, we also add a loss function $\mathcal{L}_{ifp}$ for individual fairness parity to minimize the difference in overall individual fairness between different subgroups. It can be defined as:

$$\mathcal{L}_{ifp} = (\mathbf{IF}_{(s=0)} - \mathbf{IF}_{(s=1)})^2, \tag{4.15}$$

$$\mathbf{IF}_s = \frac{\sum_{v_i \in \mathcal{V}_s} \sum_{v_j \in \mathcal{V}} \|\mathbf{Z}[i,:] - \mathbf{Z}[j,:]\|_2^2 \mathbf{W}[i,j]}{m_s}, \tag{4.16}$$

where $\mathbf{IF}_s$ represents the overall individual unfairness level in a sensitive subgroup s, $\mathcal{V}_s$ is the node set whose nodes' sensitive attributes equal to s, and m_s is the number of nonzero pairwise similarities for nodes in the node set $\mathcal{V}_s$ against all nodes in $\mathcal{V}$.

Therefore, in our proposed framework, the loss function of individual fairness can be expressed as a weighted sum of losses of overall individual fairness and individual fairness parity, weighted by hyperparameters α and β:

$$\mathcal{L}_I = \alpha \cdot \mathcal{L}_{oif} + \beta \cdot \mathcal{L}_{ifp}, \tag{4.17}$$

Eventually, the overall optimization function of BIG is composed of three loss functions, which are $\mathcal{L}_U$ to achieve the utility goal of the GNN model, $\mathcal{L}_G$ to achieve group fairness, and $\mathcal{L}_I$ to achieve robust individual fairness. It can be formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_U + \lambda \cdot \mathcal{L}_G + (1 - \lambda) \cdot \mathcal{L}_I, \tag{4.18}$$

Here, the regularization coefficient λ controls the contributions of the group fairness objective and the robust individual fairness constraint. The pipeline of the BIG's optimization is presented in Fig. 4.2.

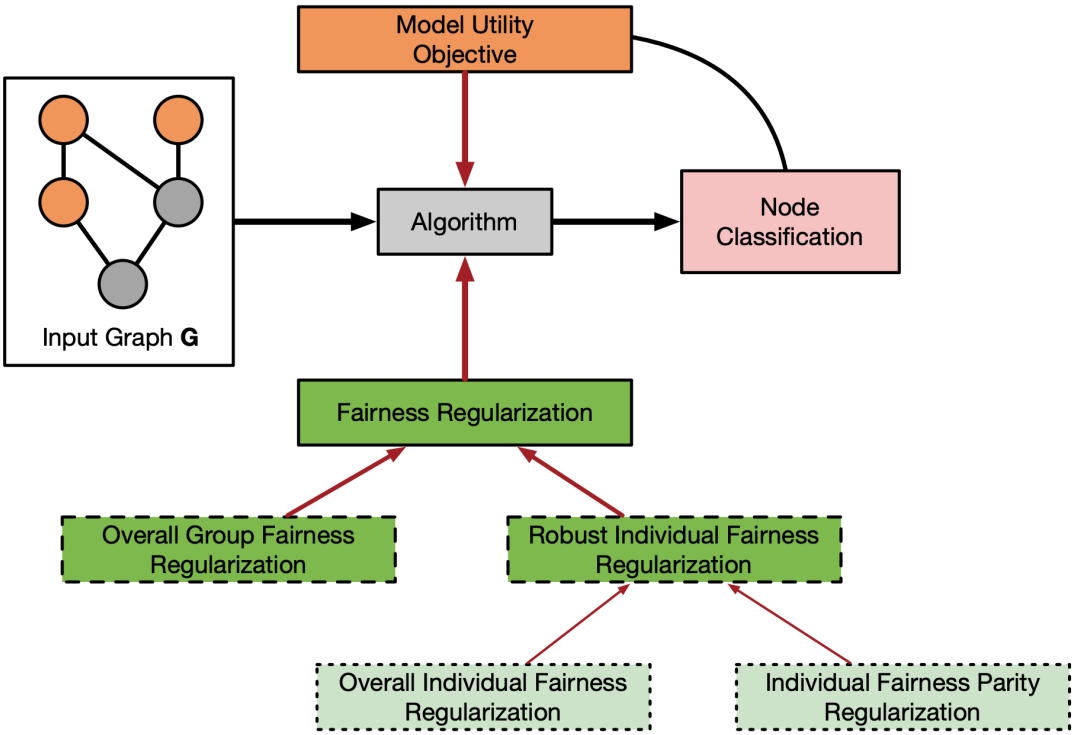


Figure 4.2: The pipeline of the BIG's optimization.

4.5 Experiments

In this section, we conduct experiments on three real-world datasets to show the effectiveness of our proposed framework BIG for binary node classification tasks. First, we introduce the datasets, compared methods, and evaluation metrics in BIG. Then particularly, we aim to answer the following essential research questions:

- **RQ1:** Can BIG achieve group fairness and individual fairness?
- **RQ2:** How well can BIG balance individual fairness, group fairness, and downstream performance objectives compared to baselines?
- **RQ3:** How does the interplay between individual fairness and group fairness, and the interplay between overall individual fairness and individual fairness parity, affect the performance of BIG?

4.5.1 Datasets

We construct three real-world datasets for our experiments, which are from the finance and demographics domain. The details of the datasets are described below. The key statistics of the datasets are shown in Table 4.1.

- **Credit defaulter**: The credit defaulter graph dataset was collected from a vital bank in Taiwan in 2005 (Yeh & Lien, 2009). The dataset contains payment data of 30,000 credit card holders and records users' attribute information, such as gender, education level, etc. Here, we treat age as the sensitive attribute, and the classification task is to predict whether a cardholder will default on the payment.

- **Census income**: The Census income graph dataset is sampled from the US Census database (Becker & Kohavi, 1996). It comprises 14,821 individuals with various features such as age, workclass, etc. We treat race as the sensitive attribute. The classification task is to predict whether a person's income exceeds $50K a year.

- **German credit**: The German credit graph dataset contains 1,000 clients' information in a German bank (Hofmann, 1994). We treat the clients' gender as the sensitive attribute, and the classification task is to predict whether the client belongs to good or bad credit.

4.5.2 Compared Methods

To validate the effectiveness of our proposed framework, we compare BIG with the following representative and state-of-art methods:

- **FairGNN** (Dai & Wang, 2021) utilizes a GNN classifier, a GCN-based sensitive attribute estimator, and an adversary to build a framework for learning fair node embeddings and classification predictions. Its core idea is to achieve group fairness in node classification tasks through adversarial learning.

Table 4.1: Statistics of datasets.

Composition	**Credit defaulter**	**Census income**	**German credit**
Nodes	30,000	14,821	1,000
Node features	13	14	27
Edges in **A**	304,754	100,483	49,940
Edges in **S**	1,687,444	1,997,641	526,982
Group ratio	11.2	3.16	2.23
Sensitive attribute	Age (≤ 25 / > 25)	Race (Black/White)	Gender (Male/Female)
Node labels	Default / No default	Income $\geq$ 50K / $<$ 50K	Good credit / Bad credit

- **NIFTY** (Agarwal et al., 2021) generates an augmented view by adding node-level, sensitive attributes, and edge-level perturbations based on the attribute information and structure of the raw graph. Then, through the defined overall objective function, its underlying GNN aims to learn node embeddings invariant to the perturbations and achieve group fairness in binary node classifications.

- **PFR** (Lahoti et al., 2019) is a mathematical optimization model and an unsupervised representation learning method that learns graph embeddings through pairwise constraints to achieve individual fairness for node classification of real-life data.

- **InFoRM** (Kang et al., 2020) defines individual fairness in graph mining based on the Lipschitz condition and deploys three mutually complementary algorithmic frameworks to mitigate the individual bias within the input graph, the mining model, and the mining results, respectively.

4.5.3 Evaluation Metrics

In this chapter, we aim to achieve three optimization objectives in the proposed framework: the model's prediction accuracy, group fairness, and individual fairness. We measure the above optimization goals through the following evaluation metrics,

respectively.

Firstly, to validate the performance of node classification, we use the widely adopted AUROC and F1-Score. Secondly, to measure the performance of BIG on group fairness, as mentioned in Section 4.3, we use Δ_{SP} and Δ_{EO} as metrics. Similarly, to measure the performance of BIG on individual fairness, we use *Inconsistency* and Δ_{IFP} for measurement. Thirdly, since BIG wants to balance group and individual fairness, we use the defined **BFS** to evaluate the effectiveness. In addition, in order to compare fairness through a more comprehensive perspective, we define **BFS**-g, which is the sum of Δ_{SP}, and Δ_{EO} and **BFS**-i, which is the sum of *Inconsistency* and Δ_{IFP}.

Table 4.2: Experimental results on three graph datasets. All entries are averages and standard deviations across five independent runs. Arrows $(\uparrow, \downarrow)$ indicate the direction of better performance. All individual fairness numbers are reported in tenths. *IC* stands for inconsistency. Best performances are shown in bold.

Credit defaulter										
Model	**AUROC**(↑)	**F1**(↑)	$\boldsymbol{\Delta_{SP}}$(↓)	$\boldsymbol{\Delta_{EO}}$(↓)	**IC**(↓)	$\boldsymbol{\Delta_{IFP}}$(↓)	**BFS-g**(↓)	**BFS-i**(↓)	**BFS**(↓)	**BFS***(↓)
InFoRM	0.68±0.00	**0.79±0.00**	0.21±0.00	0.20±0.01	**0.01±0.00**	**0.01±0.00**	0.41±0.01	**0.03±0.00**	0.41±0.01	0.44±0.01
PFR	**0.70±0.00**	**0.79±0.00**	0.16±0.00	0.15±0.01	0.23±0.00	0.17±0.00	0.31±0.01	0.40±0.01	0.35±0.01	0.71±0.01
FairGNN	0.68±0.00	0.76±0.03	0.15±0.02	0.15±0.03	0.18±0.09	0.11±0.05	0.30±0.05	0.29±0.14	0.33±0.05	0.59±0.14
NIFTY	**0.70±0.00**	**0.79±0.00**	**0.14±0.01**	0.14±0.01	0.24±0.00	0.16±0.01	0.28±0.02	0.40±0.01	0.32±0.02	0.68±0.01
BIG	0.68±0.00	0.75±0.00	**0.14±0.01**	**0.13±0.01**	0.02±0.00	0.02±0.00	**0.27±0.02**	0.04±0.00	**0.27±0.02**	**0.30±0.02**
Census income										
InFoRM	0.75±0.00	0.44±0.01	0.38±0.02	0.55±0.01	**0.01±0.00**	0.04±0.00	0.94±0.04	0.05±0.00	0.94±0.04	0.99±0.03
PFR	0.75±0.00	0.48±0.00	0.28±0.00	0.35±0.00	1.21±0.04	0.74±0.01	0.63±0.00	1.95±0.06	0.82±0.01	2.57±0.06
FairGNN	0.74±0.02	0.48±0.01	0.30±0.01	0.33±0.03	0.41±0.09	0.38±0.09	0.63±0.04	0.79±0.16	0.71±0.04	1.42±0.16
NIFTY	**0.77±0.00**	**0.51±0.00**	**0.26±0.00**	**0.31±0.00**	1.83±0.02	1.01±0.01	**0.57±0.00**	2.84±0.03	0.85±0.01	3.41±0.03
BIG	0.71±0.00	0.44±0.01	0.30±0.03	0.36±0.04	0.02±0.00	**0.01±0.00**	0.66±0.07	**0.03±0.01**	**0.67±0.06**	**0.70±0.06**
German credit										
InFoRM	0.57±0.08	0.67±0.10	0.19±0.09	0.20±0.09	0.60±0.68	0.26±0.42	0.39±0.17	0.86±1.09	0.48±0.19	1.25±1.09
PFR	0.53±0.06	0.69±0.07	0.06±0.03	0.08±0.05	0.04±0.03	0.01±0.01	0.14±0.08	0.06±0.04	0.14±0.05	0.19±0.05
FairGNN	0.63±0.04	0.67±0.14	0.10±0.12	0.11±0.16	0.16±0.11	0.06±0.04	0.22±0.28	0.22±0.15	0.24±0.27	0.43±0.26
NIFTY	**0.73±0.01**	**0.80±0.01**	0.39±0.07	0.31±0.07	0.74±0.20	0.09±0.15	0.70±0.14	0.83±0.31	0.78±0.12	1.53±0.24
BIG	0.48±0.05	0.70±0.09	**0.04±0.04**	**0.02±0.02**	**0.01±0.01**	**0.00±0.00**	**0.05±0.06**	**0.01±0.02**	**0.05±0.06**	**0.06±0.07**

4.5.4 Effectiveness of BIG

To answer **RQ1** and **RQ2**, we conduct comparative experiments to evaluate our proposed BIG through the perspectives of classification prediction performance, group fairness, individual fairness, and balanced fairness. Given the scale of the individual fairness metrics we collected, we enlarged it ten times to more intuitively

distinguish the performance of each model. **BFS**-g and **BFS**-i represent the comprehensive performance of the group and individual fairness, respectively. **BFS** and **BFS*** correspond to the Balanced Fairness Score when the γ value is 1 and 10, respectively. For all fairness metrics, the smaller the value, the better the fairness performance. All experimental results are presented as the mean and standard deviations of five independent experiments and are shown in Table 4.2. We summarize our observations as follows:

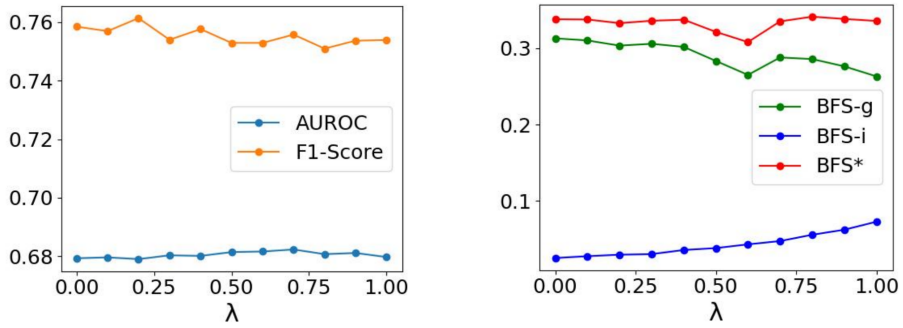
- **BIG achieves group fairness.** From the perspective of group fairness, BIG ranks first in both Δ_{SP} and Δ_{EO} compared with other baseline models in the Credit defaulter and German credit datasets. Although in the experimental results of the Census income dataset, the NIFTY model leads in Δ_{SP} and Δ_{EO}, BIG's results are close to it, showing good group fairness.

- **BIG achieves individual fairness.** From the perspective of overall individual fairness, BIG obtains its *Inconsistency* close to 0, which is significantly ahead of most other models and shows that individual fairness discrimination has been eliminated. From the perspective of individual fairness parity, BIG's Δ_{IFP} outperforms other models under the Census income and German credit datasets and ranks second with a slight gap in the Credit defaulter dataset. In summary, BIG demonstrates effective overall individual fairness and individual fairness parity in all three datasets, achieving robust individual fairness eventually.

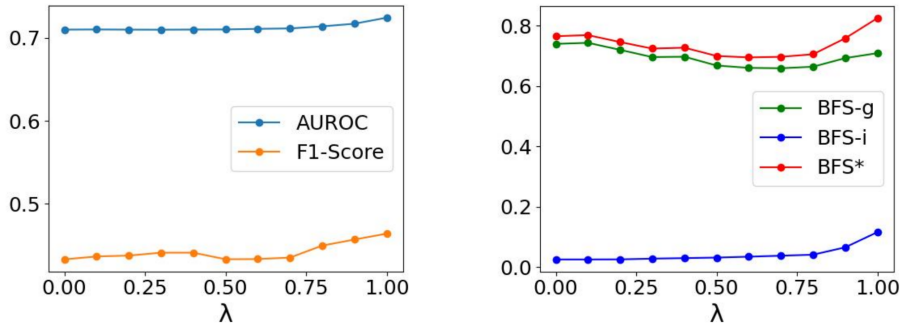
- **BIG achieves balanced fairness.** In all three datasets, BIG ranks first in the Balanced Fairness Score, regardless of whether individual fairness metrics are scaled, in other words, whether **BFS** (γ=1) or **BFS*** (γ=10). Compared with the second-best models under the three datasets, BIG achieves 31.82%, 29.29%, and 68.42% lower **BFS**, respectively, which shows that BIG considers both individual and group fairness and achieves good balanced fairness. Besides, BIG maintains relatively comparable utility performances to other

benchmark models in terms of AUROC and F1-Score. It illustrates that BIG ensures a good trade-off between the model utility and the fairness objectives.

These observations demonstrate the effectiveness of the BIG framework in achieving balanced fairness and downstream performance objectives.



(a) Credit defaulter dataset



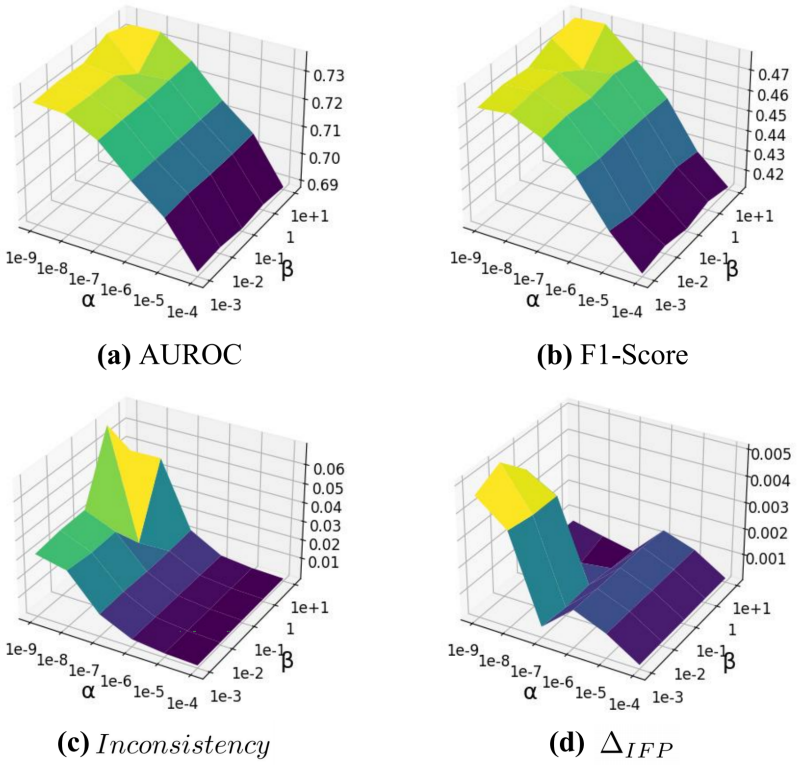
(b) Census income dataset

Figure 4.3: The effects of regularization on the performance of BIG.

4.5.5 Ablation study

To answer **RQ3**, we conduct ablation studies on two perspectives: the interplay between individual and group fairness and the interplay between overall individual fairness and its parities.

- **Interplay between individual and group fairness.** In Fig. 4.3, we compare the change patterns of different evaluation metrics as the regularization coefficient λ increases. First, AUROC and F1-Score do not fluctuate significantly with changes in λ. Second, group and individual fairness metrics show



(a) AUROC **(b)** F1-Score

(c) *Inconsistency* **(d)** Δ_{IFP}

Figure 4.4: Hyperparameter sensitivity analysis of BIG on the Census income dataset.

a clear change trend. As expected, as λ gradually becomes larger, the performance of individual fairness begins to deteriorate while the performance of group fairness gradually increases. BIG achieves the best balanced fairness in the Credit defaulter dataset when λ is between 0.55 and 0.65. In the Census income dataset, this range is $(0.6, 0.7)$. In summary, BIG's regularization coefficient λ has little impact on model utility performance but significantly impacts balancing individual and group fairness. We can find the best regularization coefficient interval for a specific dataset through experiments to obtain the best balanced fairness.

- **Interplay between overall individual fairness and its parities.** In the BIG framework, two important hyperparameters control the impact of overall individual fairness and individual fairness parity to achieve robust individual fairness. We conduct hyperparameter sensitivity experiments to investigate

their effects on achieving robust individual fairness and model utility. Specifically, we change the α value among {1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4}, and the β value among {1e-3, 1e-2, 1e-1, 1, 1e+1}. The experimental results are shown in Fig. 4.4. From Fig. 4.4(a) and 4.4(b), we observe that when $\alpha <$ 1e-7 and $\beta >$ 1e-1, the model utility on node classification can maintain good performance, and α has a more substantial impact than β. Fig. 4.4(c) shows that BIG can achieve the best *Inconsistency* when $\alpha <$ 1e-6, no matter how the β changes. From Fig. 4.4(d), we find that α and β changes cause significant Δ_{IFP} fluctuations. When $\alpha <$ 1e-6 and $\beta >$ 1e+1, Δ_{IFP} reaches the lowest value, and the best individual fairness parity is obtained.

4.6 Conclusions and Future Work

In this chapter, we propose a practical framework for balancing the conflict between group and individual fairness in GNNs. We propose a novel point of view: the essence of group and individual fairness is not a conflicted relationship and should be considered and balanced simultaneously. We accordingly propose the concept of balanced fairness and design an evaluation metric, the Balanced Fairness Score. We establish the BIG framework to achieve good group and individual fairness and maintain GNN model prediction performance. We evaluate our proposed method on three real-world datasets compared to other state-of-the-art models. Eventually, the experimental results demonstrate the effectiveness of our proposed framework.

The proposed Balanced Fairness Score (BFS) is intended to serve as a practical evaluation framework for balancing group fairness and individual fairness rather than implying that a single universally optimal fairness trade-off exists across all application scenarios. Different application scenarios may require different fairness preferences depending on regulatory, ethical, and operational considerations. Consequently, BFS should be viewed as a flexible evaluation framework that facilitates transparent comparison among competing models while allowing practitioners to

select fairness trade-offs that are appropriate for the specific regulatory, ethical, and operational requirements of their application domains.

For future work, since GNNs can perform downstream tasks among different levels, there are also fairness issues in edge-level and graph-level prediction tasks. Tasks such as link prediction, graph generation, and sub-graph optimization have had an essential impact in application fields such as healthcare and transportation, and corresponding fairness definitions and optimization methods are also required. It will be an exciting direction worth exploring.

Chapter 5

Graph-Based Machine Learning for Disease Severity Prediction and Progression Dynamics

Age-related macular degeneration (AMD) is a major cause of blindness in older adults, severely affecting vision and quality of life. Despite advances in understanding AMD, the molecular factors driving the severity of subretinal scarring (fibrosis) remain elusive, hampering the development of effective therapies. This study introduces a machine learning-based framework to predict key genes that are strongly correlated with lesion severity and to identify potential candidate genes and pathways for further biological validation to prevent subretinal fibrosis in AMD. Using an original RNA sequencing (RNA-seq) dataset from the diseased retinas of JR5558 mice, we developed a novel and specific feature engineering technique, including pathway-based dimensionality reduction and gene-based feature expansion, to enhance prediction accuracy. Two iterative experiments were conducted by leveraging Ridge and ElasticNet regression models to assess biological relevance and gene impact. The results highlight the biological significance of several key genes and demonstrate the framework's effectiveness in identifying novel potential hypotheses for future therapeutic investigation. The key findings provide valuable insights

for advancing drug discovery efforts and improving treatment strategies for AMD, with the potential to enhance patient outcomes by targeting the underlying genetic mechanisms of subretinal lesion development.

5.1 Introduction

Neovascular age-related macular degeneration (nAMD) is the leading cause of blindness in individuals aged 50 and older (Blindness, 2021). While anti-vascular endothelial growth factor (VEGF) therapies have been the gold standard for treating nAMD, long-term studies reveal that up to 70% of patients treated with anti-VEGF drugs develop subretinal fibrosis within 10 years, resulting in severe visual loss (Gillies et al., 2020)(Bloch et al., 2013). Currently, no FDA-approved treatments exist for subretinal fibrosis, making the identification of novel genetic targets and biological pathways critical for improving visual outcomes in nAMD.

The spontaneous JR5558 mouse model (Won et al., 2011), developed at the Jackson Laboratory, offers a valuable tool for studying the progression of nAMD. These mice develop subretinal fibrovascular lesions, visible as yellow mounds in fundus photographs, starting at 4 weeks and expanding until 12 weeks of age. A critical angio-fibrotic switch occurs at around 8 weeks, making the model particularly suited for examining both early neovascular changes and late-stage fibrosis (Nagai et al., 2014; Hasegawa et al., 2014; Linder et al., 2024). candidate targets for subretinal fibrosis have been validated in this model (Rossato et al., 2020). However, traditional methods of studying nAMD, which rely on observational and statistical techniques, often fall short due to the complexity and vastness of genetic data. These conventional approaches lack the precision and scalability required to identify specific genes that drive disease progression, underscoring the need for novel strategies that can address these challenges.

Machine Learning (ML) offers significant potential in the field of genomics, with its ability to analyze vast datasets and uncover intricate patterns that may not be ap-

parent through conventional analysis (Bostanci et al., 2023)(Chandrasekhar & Peddakrishna, 2023). By applying ML models to RNA sequencing (RNA-seq) data, researchers can predict lesion severity and identify key genes associated with subretinal fibrosis. Consequently, using ML to focus on disease severity in the JR5558 mouse model can further aid in high-throughput screening for novel candidate targets in subretinal fibrosis, thereby advancing drug discovery efforts. However, RNA-seq datasets can vary significantly across different applications, often facing challenges such as limited sample sizes and high dimensionality. Effectively preprocessing these datasets, selecting the most suitable models and designing appropriate prediction tasks have become critical challenges.

To address the aforementioned issues, we investigate a novel problem in this chapter: how to utilize ML models to predict lesion severity and identify potential therapeutic gene targets in subretinal fibrosis related to AMD using RNA-seq data from mice. First, we collected and organized comprehensive RNA-seq data from the retinas of JR5558 mice. We then preprocessed this original RNA-seq dataset by reducing dimensionality and expanding features before training the ML models. We employed Ridge and ElasticNet regression models for training and prediction. To verify the effectiveness of our proposed framework and further analyze the biological impact, we conducted two sets of iterative experiments on biological correlation and gene impact measurement.

The main contributions of our research can be summarized as follows:

- We collected and provided an original and comprehensive RNA-seq dataset from the retinas of JR5558 mice, facilitating further research in subretinal fibrosis.
- We applied ML models, specifically Ridge and ElasticNet regression, to predict lesion severity and identify key genes associated with subretinal fibrosis, thereby enhancing precision in genetic analysis.
- We tackled challenges related to limited sample sizes and high dimensionality

in RNA-seq data, improving data preprocessing and model selection strategies in transcriptomic research.

- We designed and conducted iterative experiments based on the original datasets we collected and produced, verifying the effectiveness and excellent performance of our proposed framework through biological impact analysis in two dimensions: biological correlation and gene impact measurement.

- Our approach identifies potential therapeutic gene targets, offering new insights into transcriptomic influences on subretinal fibrosis and advancing drug discovery efforts.

In addition to modelling disease severity from a static perspective, it is important to note that disease progression is inherently dynamic. While the proposed framework effectively identifies key genes and pathways associated with lesion severity, it does not explicitly capture the temporal evolution of these biological processes. To further address this limitation, we extend our framework by incorporating a dynamic modelling approach based on graph pseudotime analysis and neural stochastic differential equations. This extension enables us to explore the temporal progression of pathway activities and identify critical transition behaviours in disease development.

It should be noted that this chapter consists of two complementary components. The primary contribution focuses on machine learning-based prediction of disease severity and identification of biologically relevant genes from RNA-seq data. Building upon these findings, Section 5.5 presents a condensed graph-based dynamic modelling extension that illustrates how the proposed biomedical analysis may be further generalized to pathway-level disease progression modelling through Graph Pseudotime Analysis and neural stochastic differential equations. This extension is included to complement the primary biomedical study presented in this chapter rather than to constitute an independent standalone contribution.

The remainder of this chapter is structured as follows. Section 5.2 presents a

literature review of machine learning in biomedical research, RNA-seq data in machine learning and transcriptomic studies on disease severity and subretinal fibrosis. Section 5.3 illustrates our proposed framework and methodology. Section 5.4 details the dataset description, presents and discusses the experimental results and biological impact. Section 5.5 presents a dynamic modelling extension based on graph pseudotime analysis and neural stochastic differential equations. Finally, Section 5.6 summarizes the chapter's conclusions.

5.2 Related Work

5.2.1 Machine Learning in Biomedical Research

In recent years, the integration of ML into biomedical research has revolutionized the field, offering novel insights and predictive capabilities that were previously unattainable. As the volume of biomedical data continues to grow exponentially, ML provides essential methods for analyzing intricate datasets, discovering patterns and making informed predictions.

Traditional statistical methods, such as t-tests and ANOVA, have been instrumental in biomedical research but come with inherent limitations. These methods often require assumptions about the data and can struggle with the high dimensionality and complexity typically for biological datasets. Furthermore, traditional approaches may not always capture subtle patterns and interactions within the data, leading to potential oversights. In contrast, ML offers significant advantages in handling large and complex datasets. ML algorithms can uncover intricate patterns and relationships without the need for predefined models, enhancing the ability to make accurate predictions and discoveries. This flexibility is particularly beneficial in multi-omics studies, where the complexity of the data demands more robust and adaptive analytical techniques.

ML techniques have significantly advanced biomedical research, enhancing our

ability to address various critical aspects such as predicting disease outcomes, identifying biomarkers and uncovering candidate targets. In the realm of predicting disease outcomes, Khan et al. (Khan et al., 2023) and Bhatt et al. (Bhatt et al., 2023) focused on cardiovascular diseases, with Khan's employing random forest (RF) algorithms and Bhatt's utilizing a multilayer perceptron with cross-validation, both significantly enhancing early diagnosis precision. Arumugam et al. (Arumugam et al., 2023) optimized a decision tree model to predict heart disease in diabetes patients, improving diagnostic accuracy and efficiency. Similarly, Islam et al. (Islam et al., 2023) evaluated various ML models for chronic kidney disease prediction, determining that the XGBoost classifier was the most effective, thereby supporting timely intervention and treatment planning. Transitioning to the identification of biomarkers, Zhang et al. (X. Zhang, Jonassen, & Goksøyr, 2021) employed feature selection methods alongside various ML techniques, including support vector machines (SVMs), to boost diagnostic accuracy and aid in the development of targeted treatments. Mi et al. (Mi et al., 2021) introduced PermFIT, a permutation-based technique using RFs and SVMs, to identify crucial biomarkers for complex diseases. This method isolated key genetic markers, enhancing the understanding and diagnosis of intricate medical conditions. In the quest to uncover candidate targets, Pun et al. (Pun et al., 2023) utilized deep learning models to analyze large-scale omics data, identifying novel candidate targets for complex diseases by uncovering critical molecular interactions and pathways. Rafique et al. (Rafique et al., 2021) applied ensemble learning methods, such as RFs and gradient boosting machines, to predict patient responses to various cancer treatments. By analyzing clinical and molecular data, they aimed to improve the accuracy of therapeutic response predictions, ultimately enhancing patient outcomes in oncology.

With continuous advancements in this field, ML has demonstrated significant advantages in handling large-scale biological datasets and uncovering meaningful patterns. Consequently, it has become a valuable tool for processing and analyzing RNA-seq data.

5.2.2 RNA Sequencing Data in Machine Learning

RNA-seq has significantly advanced biomedical research and transcriptomics by offering a comprehensive view of gene expression. Unlike traditional profiling methods, RNA-seq quantifies transcript levels across the entire genome, providing insights into gene regulation and cellular functions (Slovin et al., 2021). This detailed analysis enables the identification of transcriptomic activity patterns associated with disease processes, which are crucial for understanding complex biological systems. Additionally, it aids in discovering potential biomarkers and candidate targets (Andrews et al., 2021). In ML applications, the ability of RNA-seq data to capture the full range of gene expression makes it an invaluable resource for developing predictive models. These models can elucidate the transcriptomic factors that influence disease severity and progression.

RNA-seq data have been instrumental in uncovering critical insights across a range of conditions, ultimately contributing to improved diagnosis, treatment and understanding of complex diseases. For instance, in oncology, RNA-seq has been used to classify cancer subtypes that predict cancer progression and treatment response (Yu et al., 2020). Additionally, in neurodegenerative diseases such as Alzheimer's, RNA-seq has revealed critical interactions between gene pairs that contribute to disease mechanisms, offering potential targets for intervention (H. Chen et al., 2019). Furthermore, in cardiovascular research, RNA-seq has provided insights into genes associated with heart failure and atrial fibrillation, aiding in the development of predictive models that enhance disease prediction and support precision medicine (Venkat et al., 2023).

Building on these advancements, ML models play a crucial role in leveraging RNA-seq data for disease research. Supervised learning models, such as SVMs and RFs, have been effectively utilized to identify biomarkers and predict treatment responses. Gupta et al. (Gupta et al., 2021) deployed RF and SVM to analyze RNA-seq datasets for identifying and validating novel transcript biomarkers associated

with hepatocellular carcinoma, focusing on improving early detection and diagnosis. Meanwhile, unsupervised learning models like hierarchical clustering have uncovered novel gene expression patterns in cancer research, offering valuable insights into underlying mechanisms. Lee et al. (Lee et al., 2020) proposed an approach that uses similarity-based hierarchical clustering to accurately analyze complex patient data and identify distinct patterns that correlate with disease progression, thereby enhancing the prediction of pathological stages in papillary renal cell carcinoma.

To enhance the efficacy of ML models applied to RNA-seq datasets, it is essential to conduct several preprocessing steps on the raw data. These steps typically include quality control measures to remove low-quality reads, serving as an initial data denoising process. In addition, one can deploy feature normalization to account for variations in sequencing depth and filtering to eliminate uninformative genes. Moreover, feature extraction methods, such as gene expression quantification and dimensionality reduction techniques such as principal component analysis (PCA) (X. Chen et al., 2020), are crucial for identifying relevant features that capture the most informative aspects of the RNA-seq data.

5.2.3 Molecular Studies on the Pathogenesis of Subretinal Fibrosis

Subretinal fibrosis is a hallmark of advanced neovascular age-related macular degeneration (nAMD) and is closely associated with poor visual outcomes (Tenbrock et al., 2022). Although anti-VEGF therapies have revolutionized the treatment of nAMD by suppressing neovascularization, their long-term effectiveness is limited. A substantial proportion of patients, despite initial responsiveness to anti-VEGF therapies, develop subretinal fibrosis within a decade, leading to irreversible vision loss (Khachigian et al., 2023). Fibrosis, characterized by the excessive accumulation of extracellular matrix (ECM) components beneath the retina, results in tissue scarring and visual impairment (Mallone et al., 2021). This underscores the need

for a deeper understanding of the molecular pathways contributing to fibrotic processes, as current treatment strategies are insufficient in halting or reversing fibrosis progression.

Recent research has uncovered critical molecular factors that influence the severity of subretinal fibrosis. Specifically, mutations in genes responsible for ECM remodeling have been implicated as key drivers of fibrosis in nAMD (Shughoury et al., 2022). Collagen and fibronectin, structural proteins of the ECM, are essential for maintaining tissue integrity and facilitating repair processes (Nita et al., 2014). However, genetic mutations in these proteins can promote aberrant ECM deposition, exacerbating fibrotic tissue development. Furthermore, the interplay between genetic predispositions and environmental factors, such as oxidative stress and chronic inflammation, accelerates the fibrotic response in the subretinal space (Kauppinen et al., 2016). Inflammatory signaling pathways, such as those mediated by cytokines and chemokines, are known to amplify the fibrotic process by enhancing the recruitment of fibroblasts and the deposition of ECM components, leading to retinal scarring. This interaction between genetic and environmental factors highlights the complexity of fibrosis and the need for multifaceted therapeutic approaches.

5.3 Methodology

5.3.1 Framework Overview

Our study, as depicted in Figure 5.1, employs a comprehensive ML-based framework designed to predict lesion severity and identify key gene targets associated with the disease using RNA-seq data from the JR5558 mouse model. This approach aims to deepen our understanding of the genetic factors influencing disease progression and to uncover potential candidate targets.

In the initial phase of our research, we collected RNA-seq data from the retinas

of 23 JR5558 mice. This data was meticulously correlated with lesion severity scores obtained from fundus photographs of these mice, where severity was quantified by measuring subretinal lesion size. This process provides us with a foundational raw RNA-seq dataset, crucial for subsequent analyses. Next, in the feature engineering stage, we tackled the challenge posed by the limited sample size and the extensive number of features. This was achieved through dimensionality reduction, specifically by organising the data according to group genes into different molecular pathways. This method allows us to focus on the most pertinent gene groups. Once these influential gene groups were identified, we expanded the dataset to concentrate on individual genes within these groups, enhancing the dataset's utility for further analysis. Following this, taking the refined dataset as input, we proceeded to the training and prediction phase using two ML models: Ridge regression and ElasticNet regression. These models were chosen for their ability to handle complex data and provide accurate predictions. Then, based on our trained models, we conduct two iterative experiments: biological correlation and influence measurement. The objective of the first experiment is to identify genes whose expression is most strongly associated with the severity of subretinal lesions in AMD, while the second experiment aims to identify target genes that, when modified, could significantly alter lesion severity, potentially revealing new treatment targets for subretinal fibrosis in AMD. Finally, we delved into a comprehensive discussion of the biological impact of our findings. This analysis is pivotal in exploring and identifying potential candidate targets, which may offer new avenues for the treatment of subretinal fibrosis.

By integrating these stages, our framework provides a robust and systematic approach to understanding the genetic influences on lesion severity, ultimately contributing valuable insights for future therapeutic interventions.

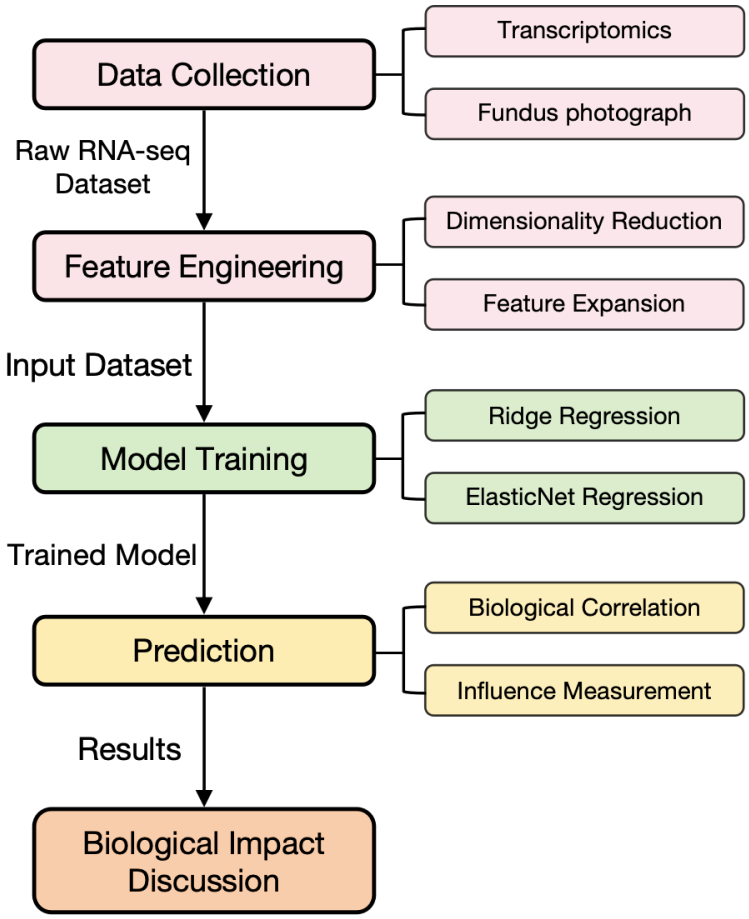


Figure 5.1: The overall framework of our study.

5.3.2 Animals

JR5558 mice were purchased from Jackson Laboratory (B6.Cg-$Crb1^{rd8}$ $Jak3^{m1J}$/Boc: JAX stock# 005558), with a genetic background of C57BL/6J. All animals were housed in the pathogen-free environment of the Animal Research Facility on a 12-light/dark cycle. The experimental procedures were conducted in accordance with the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research (`http://www.arvo.org/`). This study was approved by the Animal Ethics Committee of the University of Sydney (Project number: 2021/2013).

5.3.3 Data Collection

The study utilizes bulk RNA-seq data from the retinas of JR5888 mice, a well-established model for studying nAMD. Total RNA from twenty-three retinas of 8-week-old male JR5558 mice was extracted using GenEluteTM Single Cell RNA Purification Kit (Sigma Aldrich, RNB300). The library preparation, quality control and sequencing were commercially contracted to Novogene (`https://www.novogene.com/`). The RNA-seq dataset includes expression levels of 56,748 genes, quantified as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). The FPKM values provide a normalized measure of gene expression, accounting for gene length and sequencing depth, allowing for accurate comparisons across samples. Out of the 56,748 genes, 24,888 have corresponding Entrez IDs, which were used for subsequent pathway analysis. In addition to the RNA-seq data, fundus photographs were also taken to analyze the disease severity. In brief, mice were anesthetised by intraperitoneal injection of ketamine (48 mg/kg, Troy Laboratories, Australia) and medetomidine (0.6 mg/kg, Troy Laboratories, Australia). Mice pupils were dilated with 0.5% Tropicamide. Fundus photographs were performed with MICRON IV Retinal Imaging Microscope (Phoenix Technology Group, USA) with optic nerve locating at the center of the image. The total area of subretinal lesions on fundus images was then independently quantified by two investigators using FIJI ImageJ

software (`https://imagej.net/software/fiji/downloads`, National Institutes of Health, USA). Figure 5.2 illustrates the key steps of the quantification method. An imageJ macro was developed accordingly to ensure consistent output. Of note, lesions that either totally or partially fell into the circle with a radius of 283 μm from the optic nerve were included in the analysis. Lesion severity was quantified based on the percentage of subretinal lesion area observed in fundus photographs of the mouse retina. The lesion area serves as the primary outcome variable for predicting the severity of fibrosis in this study.

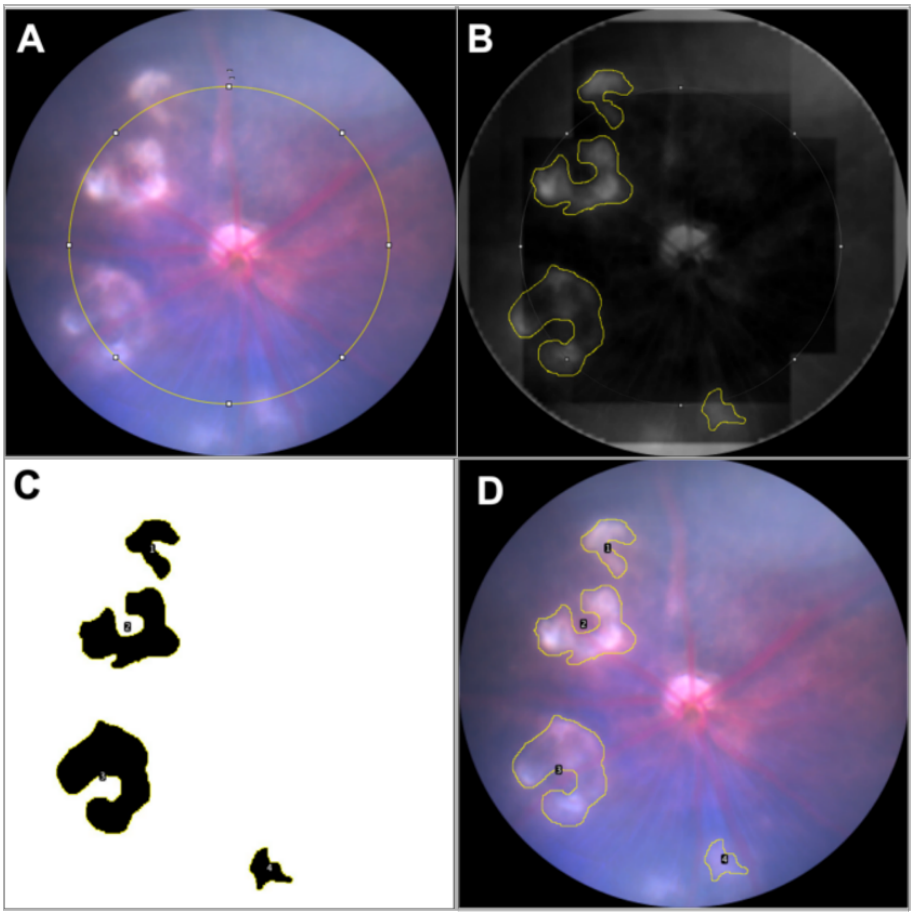


Figure 5.2: Key steps of the quantification method.

A: Draw a circle with a radius of 283 μm centered on the optic nerve. B: Convert the image to an 8-bit grayscale, remove the background and mark with freehand tools. C: Adjust the threshold to precisely select the lesion area. D: Verify in the original image that all lesions have been selected.

5.3.4 Feature Engineering

5.3.4.1 Dimensionality Reduction

Dimensionality reduction is a crucial technique in the analysis of RNA-seq data, particularly when investigating the influence of genes on disease severity. In relation to our research on the JR5558 mouse model, dimensionality reduction serves to strike a balance between model performance and computational efficiency. High-dimensional data can lead to overfitting, where the model becomes too tailored to the training data of small size, resulting in poor generalisation to unseen data. Conversely, reducing the dimensionality excessively may lead to the loss of vital biological information necessary for understanding disease progression.

Several methods are available for dimensionality reduction, including feature extraction techniques like PCA and feature selection methods such as Least Absolute Shrinkage and Selection Operator (LASSO) and RF. For our study, feature selection is particularly advantageous, as it retains the interpretability of the model. This is essential for identifying potential candidate targets, as we aim to elucidate the relationship between individual genes and the severity of subretinal fibrosis.

Incorporating domain knowledge is essential for effective feature selection from RNA-seq data, as it helps identify relevant genes that influence disease severity. Without this understanding, ML models may miss important non-linear relationships, potentially excluding critical genes. To maintain robust model performance, feature selection should be performed solely on the training dataset to prevent data leakage, which could skew performance metrics. Additionally, ensuring consistent variable types across training and testing datasets is crucial for achieving reliable predictions and generalizability.

In our study, we face a challenge common to many biological datasets: a limited number of samples from JR5558 mice, yet a vast array of gene features for each sample. To address this, we developed an approach to effectively reduce the dimensionality of the RNA-seq data by grouping genes into canonical molecular pathways. A web

scraper was developed in R to extract detailed biological pathway information related to Mus musculus from the KEGG database (`https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=mmu`). The data extracted included pathway maps, which were organized into a data frame for subsequent dimensional reduction. This step was crucial to managing the high dimensionality of the RNA-seq data by focusing on relevant pathways rather than individual genes, thus reducing noise and improving the model's ability to identify significant correlations. By grouping gene features and averaging expression values within each group, we were able to reduce the dataset's dimensionality from over 24,888 features to 343, thereby streamlining the dataset for more effective prediction and analysis.

5.3.4.2 Feature Expansion

Feature expansion is a technique used to enhance the predictive capabilities of a model by increasing the number of relevant features. It involves transforming existing data to uncover additional insights that may not be immediately apparent. This method can be particularly useful in complex biological datasets, allowing for a more comprehensive analysis by incorporating diverse aspects of the data.

Pathway-based dimensionality reduction allows for effective predictive analysis on gene-group datasets with reduced dimensionality. Once we identify the key gene groups, we can expand their original gene features to enhance training and prediction focused on these crucial genes. Specifically, genes from the identified key biological pathways are extracted and their FPKM values from each sample are used as features for the second round of data processing. This gene-based feature expansion allows the model to incorporate detailed expression information for genes that are potentially involved in fibrosis, enhancing its predictive power.

5.4 Experiments

5.4.1 Dataset Description

The dataset for this study comprises bulk RNA-seq data from the retinas of JR5888 mice and corresponding lesion area measurements. The RNA-seq data included expression levels for 56,748 genes, with 24,888 genes associated with Entrez IDs. The lesion area data were derived from fundus photographs, where the extent of subretinal fibrosis was quantified as a percentage of the total retinal area. This dataset was divided into training and validation sets to evaluate the performance of the ML models in predicting key genes associated with disease severity.

5.4.2 Prediction and Results

In this study, we conducted two prediction tasks: 1. biological significance and 2. influence measurement, each involving two rounds of iterative experiments. Fig. 5.3 illustrates the logic of our experimental approach. In the first round, we employed a dataset with gene groups as features, utilizing dimensionality reduction through a gene pathway-based method to perform predictive tasks. Based on these initial results, we identified the top-performing gene groups as candidates for further analysis. In the second round, we expanded our focus by using the expression values of the original genes within these selected candidates. This allowed us to conduct more detailed prediction tasks and perform biological analysis targeting individual genes, thereby enhancing the depth and accuracy of our findings.

In this study, we conducted two prediction tasks: 1. biological significance and 2. influence measurement, each involving two rounds of iterative experiments. Figure 5.3 illustrates the logic of our experimental approach. In the first round, we employed a dataset with gene groups as features, utilizing dimensionality reduction through a gene pathway-based method to perform predictive tasks. Based on these initial results, we identified the top-performing gene groups as candidates for further

analysis. In the second round, we expanded our focus by using the expression values of the original genes within these selected candidates. This allows us to conduct more detailed prediction tasks and perform biological analysis targeting individual genes, thereby enhancing the depth and accuracy of our findings.

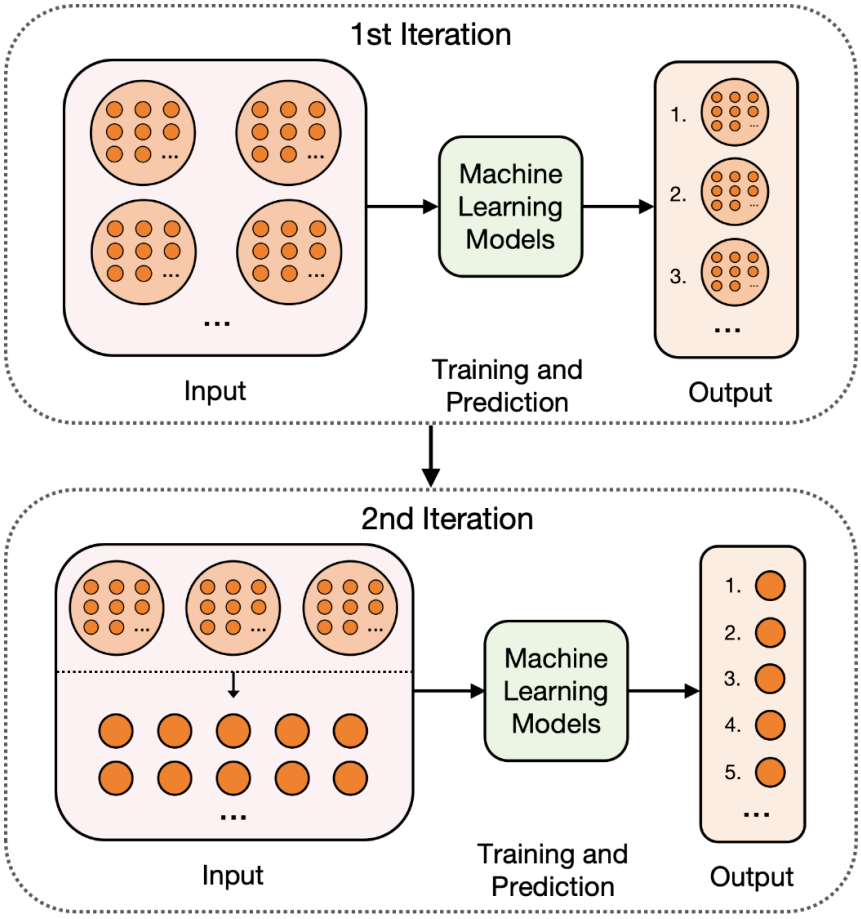


Figure 5.3: The logic of our iterative experiments.

5.4.2.1 Biological Correlation

The objective of the first task is to predict the genes most strongly correlated with the severity of disease progression in AMD.

In the first round of experiments, we used Ridge regression and ElasticNet regression models to train and predict the gene group dataset following dimensionality reduction, identifying the overlapping most influential gene groups from both models. We identified nine recurring candidates from the top ten most influential gene

Table 5.1: Top 7 most important genes selected by our ML models in biological correlation experiments with the greatest association to the disease.

Ranking	**Influential Gene Entrez ID**	**Gene Name and Description**
1	16177	interleukin 1 receptor, type I
2	18796	phospholipase C, beta 2
3	12259	complement component 1, q subcomponent, alpha polypeptide
4	320302	glycosyltransferase 28 domain containing 2
5	15212	hexosaminidase B
6	18018	nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 1
7	20848	signal transducer and activator of transcription 3

groups produced by both models. In the second round of experiments, we expanded the gene expression features within these influential gene groups to further identify the most significant genes. The seven selected most influential genes, which have the greatest association with the disease in biological correlation experiments, are listed in Table 5.1.

5.4.2.2 Influence Measurement

The purpose of the second task is to identify target genes that, when manipulated, could significantly modulate the severity of subretinal fibrosis, potentially leading to more effective therapeutic strategies.

In the first round, we adjusted each gene group feature by decreasing its expression value to 50% and increasing it to 200%, respectively. We then applied the trained model to predict lesion severity based on these modified gene group expressions and recorded the changes in predicted severity. We calculated the differences in lesion severity before and after the adjustments to each gene group's expression. This allows us to identify the gene groups whose modifications resulted in the most significant changes in lesion severity. Here, we counted and ranked the changes in lesion severity into two categories: aggravation and alleviation. Therefore, we iden-

tified the most significant gene groups that either worsened or alleviated the disease as candidates. By the end of the first round of experiments, we obtained two rankings for the gene groups: 50% manipulation that alleviated the disease, and 200% manipulation that caused the disease to worsen.

Table 5.2: Top 10 genes selected by our ML models for their significant impact on disease improvement when gene expression values were reduced by 50%.

Ranking	**Influential Gene Entrez ID**	**Gene Name and Description**
1	66357	oligosaccharyltransferase complex subunit (non-catalytic)
2	26416	mitogen-activated protein kinase 14
3	12260	complement component 1, q subcomponent, beta polypeptide
4	16179	interleukin-1 receptor-associated kinase 1
5	320302	glycosyltransferase 28 domain containing 2
6	12262	complement component 1, q subcomponent, C chain
7	224530	acetyl-Coenzyme A acetyltransferase 3
8	14676	guanine nucleotide binding protein, alpha 15
9	16176	interleukin 1 beta
10	12503	CD247 antigen

In the second round of experiments, we performed feature expansion on the two ranked gene group lists from the first round and obtained two corresponding datasets with individual gene expression values as features. For each dataset, we performed the same manipulations as in the previous round and counted the rankings that caused the same disease impact. This time, we identified the most influential individual genes. For example, for the gene groups that resulted in 50% manipulation causing the disease alleviation in the previous round, we continued to apply the same 50% manipulation after expanding the gene dataset and ranked the genes that had the same impact on the disease, which was improving. This process allowed us to identify the genes that most significantly affected lesion severity. Tables 5.2 and 5.3 have shown the final results of the influence measurement experiments, respectively.

Table 5.3: Top 10 genes selected by our ML models for their significant impact on disease exacerbation when gene expression values were increased to 200%.

Ranking	Influential Gene Entrez ID	Gene Name and Description
1	18798	phospholipase C, beta 4
2	12260	complement component 1, q subcomponent, beta polypeptide
3	12259	complement component 1, q subcomponent, alpha polypeptide
4	12322	calcium/calmodulin-dependent protein kinase II alpha
5	12262	complement component 1, q subcomponent, C chain
6	19091	protein kinase, cGMP-dependent, type I
7	106759	toll-like receptor adaptor molecule 1
8	20293	chemokine (C-C motif) ligand 12
9	224530	acetyl-Coenzyme A acetyltransferase 3
10	12789	cyclic nucleotide gated channel alpha 2

5.4.3 Biological Impact Discussion

The integration of ML models with RNA-seq data offers a robust approach for identifying transcriptomic factors linked to disease severity in AMD. This study specifically aimed at predicting genes associated with the severity of subretinal lesions in a mouse model of AMD. By employing dimensionality reduction and feature expansion techniques, our model successfully identified genes within specific pathways that their related proteins and signaling pathways are likely contributors to subretinal fibrosis that may be considered as candidate targets for further biological validation. These findings enhance our understanding of the molecular mechanisms underlying subretinal fibrosis and present new opportunities for therapeutic interventions.

It is remarkable that two different tasks with different algorithms identified several common key genes/proteins or targeting same pathway. We first compared the top 10 target genes in table II and table III and identified two proteins is positively correlated to the lesion severity. They are Complement component 1, q subcomponent (C1q) and Acetyl-Coenzyme A acetyltransferase 3 (ACAT3). We further

compared the results from task 1 and task 2 and further two more common proteins has been identified. They are Phospholipase C (PLC), and Glycosyltransferase 28. All four common proteins were strongly implicated in both the progression of subretinal fibrosis and as potential candidate targets (Tables 5.1, 5.2 and 5.3). Each of these proteins operates within distinct yet interconnected biological processes, highlighting a complex interaction between inflammation, immune responses, cellular signaling and metabolic regulation that drives fibrotic changes in the retina.

The C1q proteins are key components of the classical complement pathway which initiates immune surveillance, inflammation and the clearance of apoptotic cells (Cho, 2019) (Galvan et al., 2012). Dysregulation of this pathway, particularly sustained activation of C1q, has been closely linked to the progression of AMD (Y. Ma et al., 2022). Ongoing C1q activation in subretinal fibrosis fuels chronic inflammation, promoting tissue damage and extracellular matrix remodeling—both hallmarks of fibrosis. The accumulation of matrix components in the subretinal space results in tissue thickening and scarring, which contribute to irreversible loss of vision in advanced AMD. It is worth noting that we also identified several inflammation-related molecules, including Interleukin 1 receptor, type I (IL1R1, Table 5.1), Signal transducer and activator of transcription 3 (STAT3, Table 5.1), Interleukin-1 receptor-associated kinase 1 (IRAK1, Table 5.2), Interleukin 1 beta (IL1B, Table 5.2), Toll-like receptor adaptor molecule 1 (TICAM1, Table 5.3) and Chemokine (C-C motif) ligand 12 (CCL12, Table 5.3). This suggests a strong positive correlation between the activation of inflammatory pathways and the severity of the subretinal lesion.

PLC is critical for intracellular signal transduction. This signaling pathway is also essential for regulating cellular responses to inflammatory stimuli, proliferation and apoptosis (Y.-N. Wu et al., 2023). PLC may intensify complement activation in subretinal fibrosis by amplifying pro-inflammatory signals (Zhu et al., 2018). This interaction between PLC and the complement system may create a self-perpetuating cycle of inflammation that drives tissue remodeling and fibrosis in the retina.

Acetyl-Coenzyme A acetyltransferase (ACAT) plays a key role in acetyl-CoA metabolism by influencing cholesterol and production of ketone bodies (Z. Ma et al., 2023). Its involvement in lipid metabolism is particularly relevant to retinal diseases, where metabolic dysregulation often exacerbates inflammation and tissue damage (Ana et al., 2023). Altered ACAT activity could disrupt lipid homeostasis, promoting cellular stress and inflammation in retinal cells, which may accelerate the development of subretinal fibrosis by further stimulating inflammatory and fibrotic processes.

Glycosyltransferase is an enzyme responsible for glycosylation, the addition of sugar moieties to proteins and lipids. This modification impacts protein folding, stability and interactions, all of which are crucial for maintaining cellular function (Nagare et al., 2021). Aberrant glycosylation has been associated with fibrosis across various tissues (Loaeza-Reyes et al., 2021). In subretinal fibrosis, Glycosyltransferase may influence key protein modifications involved in inflammation and extracellular matrix formation, potentially exacerbating tissue scarring and the progression of fibrosis.

Together, these four proteins, C1q, PLC, ACAT3 and Glycosyltransferase, likely form an interconnected network that sustains chronic inflammation and metabolic dysfunction in the retina. Their combined activity promotes extracellular matrix deposition and tissue remodeling, driving the progression of subretinal fibrosis. Understanding their precise roles and interactions could provide critical insights into therapeutic strategies for AMD. Targeting these proteins or their signaling pathways may offer effective ways to reduce inflammation, slow fibrotic changes and prevent vision loss in patients with AMD.

These results underscore the potential of combining ML with molecular imaging techniques to enhance our understanding of fibrotic diseases and improve patient outcomes. Future research could explore the application of this approach to other fibrotic conditions and assess its potential for personalising treatment strategies based on individual genetic profiles.

5.5 Dynamic Modelling Extension via Graph Pseudotime Analysis and Neural SDEs

The preceding sections primarily focus on predicting lesion severity and identifying influential genes from a static transcriptomic perspective. However, disease progression is intrinsically dynamic, and static prediction alone cannot fully capture how pathway activities evolve during retinal degeneration. To complement the above analysis, we further explore a dynamic modelling extension that aims to characterize disease progression at the pathway level.

5.5.1 Motivation

A major challenge in transcriptomic studies of disease progression is that longitudinal observations from the same subject are often unavailable. In our setting, each RNA-seq sample provides a snapshot of one disease state, but not a directly observed temporal sequence. This makes it difficult to model progression dynamics using conventional time-series methods. To address this issue, we introduce a graph-based pseudotime framework that infers an intrinsic ordering among pathway graphs and then models the resulting pathway trajectories via neural stochastic differential equations (SDEs).

5.5.2 Pathway Graph Construction

To move from gene-level prediction to pathway-level progression modelling, we first organize genes into canonical molecular pathways using biological prior knowledge. Specifically, genes with Entrez identifiers are grouped into KEGG pathways, and each sample is represented as a pathway graph. In this graph, each node corresponds to one molecular pathway, while the node features are derived from the gene expression profiles associated with that pathway.

Edges are constructed between pathways when gene overlap exists, and edge

weights are used to characterize the similarity between pathway-level features. In this way, each mouse retina is represented by a graph object that captures both pathway composition and pathway relationships. Compared with the static regression framework above, this graph representation provides a more structured basis for modelling pathway interactions and disease progression.

5.5.3 Graph Pseudotime Analysis

After constructing pathway graphs for all samples, we apply Graph Pseudotime Analysis (GPA) to infer an approximate disease trajectory from cross-sectional observations. The purpose of GPA is to recover an intrinsic ordering of the graph samples, so that unordered transcriptomic snapshots can be rearranged into a pseudotemporal progression.

Concretely, each graph is first mapped to a low-dimensional representation by combining graph feature information and graph structural information. A graph-level embedding is then obtained through pooling and dimensionality reduction. Based on these representations, we construct a neighbourhood graph and further estimate a minimum spanning tree to preserve the global structure of the sample manifold. By selecting low-severity and high-severity samples as endpoints, shortest paths on this graph can be used to estimate putative disease trajectories.

This procedure allows us to move beyond static severity prediction and to study progression patterns at the population level, even when true longitudinal data are unavailable.

5.5.4 Neural SDE for Pathway Dynamics

Given the inferred pseudotime trajectories, we then model the evolution of pathway-level features using neural stochastic differential equations. Let $\mathbf{x}(t)$ denote the feature state of one pathway along the inferred disease trajectory. We assume that

its dynamics can be described by

$$\frac{\partial \mathbf{x}(t)}{\partial t} = \psi_{\boldsymbol{\theta}}(\mathbf{x}(t), t)\, dt + \xi_{\boldsymbol{\phi}}(\mathbf{x}(t), t)\, d\mathbf{B}(t), \tag{5.1}$$

where $\psi_{\boldsymbol{\theta}}$ denotes the drift term that captures deterministic progression trends, $\xi_{\boldsymbol{\phi}}$ denotes the diffusion term that characterizes stochastic fluctuations, and $d\mathbf{B}(t)$ is the standard Wiener process.

This formulation is particularly suitable for biological progression modelling because it captures both systematic disease evolution and random perturbations in pathway activities. In contrast to a purely deterministic model, the neural SDE framework provides a more flexible description of transcriptomic changes during retinal degeneration.

5.5.5 Pathway Stability

One advantage of the neural SDE formulation is that it enables the quantitative analysis of pathway stability. Intuitively, pathways with relatively stable diffusion behaviour are less sensitive to perturbation, whereas pathways with large stochastic variation may indicate unstable biological regulation during disease progression.

Following this idea, pathway stability can be quantified by the temporal variation of the diffusion term:

$$\mathrm{PS}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} \left\| \xi_{\phi}(\mathbf{x}(t+1), t+1) - \xi_{\phi}(\mathbf{x}(t), t) \right\|^2, \tag{5.2}$$

where T denotes the total number of pseudotime steps. Smaller values of $\mathrm{PS}(\mathbf{x})$ indicate more stable pathways, whereas larger values suggest stronger instability. This analysis provides a new perspective on identifying which pathways remain relatively robust during progression and which pathways are more vulnerable to dynamic perturbation.

5.5.6 Disease Bifurcation Points

Beyond pathway stability, the dynamic framework also allows us to investigate critical transitions during disease progression. In particular, we are interested in identifying bifurcation points, namely transition stages beyond which the system may enter a new and potentially irreversible disease state.

In this chapter, the identified bifurcation points are interpreted as *model-inferred transition points* in the reconstructed disease progression trajectory. Operationally, they correspond to stages where pathway behaviour exhibits substantial qualitative changes, such as loss of stability, transition to a new dynamic state, or increased stochastic fluctuation. From a biomedical perspective, these inferred transitions may highlight potentially important stages in disease progression that warrant further biological investigation. However, their biological and clinical significance should be interpreted with caution, as experimental and longitudinal validation would be required to determine whether they correspond to irreversible pathological changes or clinically meaningful intervention windows.

5.5.7 Discussion

This dynamic modelling extension complements the static machine learning framework developed earlier in this chapter. While the static framework identifies key genes and pathways associated with lesion severity, the graph pseudotime and neural SDE extension provides an additional view of how pathway activities may evolve over the course of disease progression.

It should be emphasized that this part serves as an extension rather than the main focus of the chapter. Nevertheless, it demonstrates the potential of integrating graph-based transcriptomic representations with stochastic dynamical systems to study retinal disease progression from a more mechanistic perspective. Future work may further refine this framework by incorporating pathway interactions, validating inferred trajectories with richer longitudinal data, and integrating multi-omics

measurements for a more comprehensive characterization of disease dynamics.

5.6 Conclusions

In this chapter, we have presented a comprehensive approach to addressing the challenges of predicting lesion severity in nAMD using an ML-based framework. We introduced a unique RNA-seq dataset derived from the JR5558 mouse model, which provides a valuable resource for further research in subretinal fibrosis. We successfully addressed issues of limited sample sizes and high dimensionality typical of RNA-seq data by employing dimensionality reduction and feature expansion techniques. Utilizing Ridge and ElasticNet regression models, our iterative experiments confirmed the effectiveness of our framework, highlighting its potential in identifying critical genetic targets linked to subretinal fibrosis.

Building upon this static modelling framework, we further explored the dynamic nature of disease progression. By incorporating graph pseudotime analysis and neural stochastic differential equations, we introduced a novel approach for modelling the temporal evolution of pathway activities. This extension enables the characterization of pathway stability and the identification of potential bifurcation points, providing deeper insights into critical transitions during disease development.

The findings of this chapter contribute to both methodological development and biological understanding. From a computational perspective, the proposed framework demonstrates the effectiveness of combining machine learning with biologically informed feature engineering. From a biomedical perspective, the identified genes and pathways offer valuable insights into the molecular mechanisms underlying subretinal fibrosis and suggest potential candidate targets for nAMD.

Overall, this work establishes a unified framework that bridges static prediction and dynamic analysis of disease processes. It provides a foundation for future research on integrating multi-omics data and developing more advanced models to better understand disease mechanisms and support precision medicine.

Chapter 6

Conclusions

6.1 Summary of the Thesis

In recent years, the increasing availability of complex and structured data has posed significant challenges for traditional machine learning methods. Many real-world systems, such as financial transaction networks and biological systems, are inherently relational and involve intricate dependencies among entities. Conventional approaches that treat data as independent samples often fail to fully capture these relationships, limiting their effectiveness in modeling complex phenomena. Graph-based learning, particularly graph neural networks (GNNs), has emerged as a powerful paradigm for representing and analyzing such structured data. This thesis investigates the development and application of graph-based and machine learning frameworks for modeling complex real-world systems, with a focus on financial applications, fairness-aware learning, and biomedical data analysis.

First, this thesis addresses the problem of credit card fraud detection in real-world transactional datasets. A key challenge in this domain lies in the presence of missing values and the underutilization of raw transactional information. To overcome these limitations, a heterogeneous graph construction method, namely the Credit-card Heterogeneous Transactional graph (CHET-graph), is proposed to effectively represent transaction data. Building upon this graph representation, a

feature importance-based edge weighting strategy is introduced to enhance interpretability. These components are integrated into a graph neural network framework, FIW-GNN, which demonstrates improved robustness and superior detection performance on real-world datasets.

Second, this thesis explores fairness issues in GNNs. Existing research often treats group fairness and individual fairness as conflicting objectives, leading to solutions that optimize only one aspect of fairness. This thesis challenges this assumption by proposing a new perspective that these two forms of fairness can be balanced. A novel concept of balanced fairness and a corresponding evaluation metric, the Balanced Fairness Score, are introduced. Based on these ideas, the BIG framework is developed to jointly optimize group and individual fairness while maintaining predictive performance. Experimental results demonstrate that the proposed framework achieves a desirable balance between fairness and accuracy in graph-based learning tasks.

Third, this thesis investigates the application of machine learning methods in biomedical data analysis, particularly in understanding the genetic mechanisms of age-related macular degeneration (AMD). A machine learning-based framework is developed to predict lesion severity and identify key genes associated with subretinal fibrosis using RNA sequencing data. By incorporating pathway-based dimensionality reduction and gene-level feature expansion, the proposed approach effectively addresses challenges such as high dimensionality and limited sample sizes. The results provide biologically meaningful insights and identify potential therapeutic targets, demonstrating the value of integrating machine learning into genomic research.

Building upon this primary biomedical study, the thesis further presents a condensed graph-based dynamic modelling extension based on Graph-level Pseudotime Analysis (GPA) and neural stochastic differential equations. This extension is included to illustrate how the proposed framework may be further generalized to investigate pathway-level disease progression and dynamic biological processes. It is presented as a complementary methodological extension to the primary contri-

bution on disease severity prediction and biologically relevant gene identification, rather than as an independent standalone contribution.

Overall, this thesis demonstrates that graph-based and machine learning approaches provide a flexible and powerful framework for modeling complex systems across different domains. By integrating representation learning, fairness considerations, and domain-specific applications, the proposed methods contribute to both methodological advancements and practical applications in financial and biomedical fields.

6.2 Key Contributions

This thesis makes several important contributions to the field of graph-based learning and its applications in real-world domains.

First, this thesis proposes novel graph-based modeling frameworks for handling complex and noisy real-world data. In the context of financial systems, a heterogeneous graph construction method is developed to effectively represent transactional data with missing values. By incorporating feature importance into edge weighting, the proposed approach enhances both model interpretability and predictive performance, demonstrating the practical advantages of graph-based representations in fraud detection tasks.

Second, this thesis advances the understanding of fairness in graph neural networks. It challenges the widely held assumption that group fairness and individual fairness are inherently conflicting objectives. By introducing the concept of balanced fairness and the Balanced Fairness Score, this work provides a unified perspective for evaluating fairness in graph-based models. The proposed BIG framework further demonstrates that it is possible to simultaneously optimize multiple fairness objectives while maintaining competitive predictive performance.

Third, this thesis contributes to the application of machine learning in biomedical data analysis. A novel framework is developed for analyzing RNA sequencing

data to predict disease severity and identify key genes associated with subretinal fibrosis. By addressing challenges such as high dimensionality and limited sample sizes through tailored feature engineering strategies, this work demonstrates the potential of machine learning methods in uncovering biologically meaningful insights and supporting therapeutic discovery.

Finally, this thesis extends graph-based learning to the modeling of biological dynamics. By introducing Graph Pseudotime Analysis and integrating neural stochastic differential equations, this work provides a new approach for analyzing disease progression at the pathway level. This contribution enables the identification of critical transitions, pathway stability, and disease bifurcation points, offering a deeper understanding of complex biological processes.

Overall, these contributions demonstrate the effectiveness and versatility of graph-based and machine learning approaches in addressing challenges across multiple domains, including finance, fairness-aware learning, and biomedical research.

6.3 Limitations and Future Work

Despite the contributions of this thesis, several limitations should be acknowledged.

First, the proposed methods are primarily evaluated on specific datasets within each domain, including financial transaction data and biomedical datasets derived from experimental studies. While the results demonstrate strong performance in these settings, the generalizability of the proposed approaches to other datasets and application scenarios may require further validation.

Second, although the proposed frameworks improve interpretability compared to many existing approaches, particularly through feature importance-based weighting and structured modeling, there remains a gap between model outputs and fully interpretable domain knowledge. In high-stakes applications such as finance and healthcare, further efforts are needed to enhance transparency and trustworthiness.

Third, the modeling frameworks presented in this thesis are mainly developed

under controlled experimental conditions. Real-world deployment of such models may introduce additional challenges, including data heterogeneity, distribution shifts, and computational constraints, which are not fully addressed in this work.

Finally, while the integration of graph-based learning and dynamic modeling provides valuable insights into complex systems, the scalability of these methods to large-scale datasets and real-time applications remains an open challenge.

These limitations highlight opportunities for further research and development to enhance the robustness, interpretability, and applicability of graph-based learning methods in real-world scenarios.

Building upon the findings of this thesis, several promising directions for future research can be identified. One important direction is the further development of fairness-aware graph learning methods. While this thesis demonstrates that group and individual fairness can be balanced, future work may explore fairness in more complex settings, such as dynamic graphs, edge-level and graph-level prediction tasks, and multi-objective optimization under real-world constraints.

Another promising direction is the extension of graph-based models to dynamic and temporal data. Many real-world systems, particularly in biomedical and financial domains, evolve over time. Developing more advanced temporal graph learning frameworks and integrating them with dynamic modeling techniques, such as stochastic differential equations, could provide deeper insights into system evolution and improve predictive capabilities.

In the biomedical domain, future research may focus on integrating multi-modal and multi-omics data, including genomics, transcriptomics, and proteomics. Combining these heterogeneous data sources with graph-based learning frameworks has the potential to provide a more comprehensive understanding of disease mechanisms and to accelerate the discovery of therapeutic targets.

In addition, improving the scalability and efficiency of graph-based models remains an important challenge. As datasets continue to grow in size and complexity, developing computationally efficient algorithms and leveraging distributed or paral-

lel computing techniques will be critical for enabling large-scale applications.

Finally, bridging the gap between research and real-world deployment is an essential direction. Future work may focus on translating graph-based learning models into practical systems that can be applied in industry and clinical settings, with an emphasis on robustness, interpretability, and user trust.

These directions highlight the broad potential of graph-based and machine learning approaches for advancing both theoretical research and practical applications in complex real-world systems.

References

Agarwal, C., Lakkaraju, H., & Zitnik, M. (2021). Towards a unified framework for fair and stable graph representation learning. In *Uncertainty in ai* (pp. 2114–2124).

Alharbi, F., Vakanski, A., Zhang, B., Elbashir, M. K., & Mohammed, M. (2025). Comparative analysis of multi-omics integration using graph neural networks for cancer classification. *IEEE Access*.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., ... Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, *12*(19), 9637.

Ana, R. d., Gliszczyńska, A., Sanchez-Lopez, E., Garcia, M. L., Krambeck, K., Kovacevic, A., & Souto, E. B. (2023). Precision Medicines for Retinal Lipid Metabolism-Related Pathologies. *Journal of Personalized Medicine*, *13*(4), 635.

Andrews, T. S., Kiselev, V. Y., McCarthy, D., & Hemberg, M. (2021). Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nature Protocols*, *16*(1), 1–9.

Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using machine learning algorithms. *Materials Today: Proceedings*, *80*, 3682–3685.

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In *International conference on computing networking and informatics (iccni)* (pp. 1–9).

Becker, B., & Kohavi, R. (1996). *Adult.* UCI Machine Learning Repository. (https://doi.org/10.24432/C5XW20)

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3–44.

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88.

Bing, R., Yuan, G., Zhu, M., Meng, F., Ma, H., & Qiao, S. (2023). Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications. *Artificial Intelligence Review*, *56*(8), 8003–8042.

Binns, R. (2020). On the apparent conflict between individual and group fairness. In *Acm facct* (pp. 514–524).

Blindness, G. (2021). Vision Impairment C, Vision Loss Expert Group of the Global Burden of Disease S. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*, *9*(2), e144–e160.

Bloch, S. B., Lund-Andersen, H., Sander, B., & Larsen, M. (2013). Subfoveal fibrosis in eyes with neovascular age-related macular degeneration treated with intravitreal ranibizumab. *American Journal of Ophthalmology*, *156*(1), 116–124.

Bostanci, E., Kocak, E., Unal, M., Guzel, M. S., Acici, K., & Asuroglu, T. (2023). Machine learning analysis of RNA-seq data for diagnostic and prognostic prediction of colon cancer. *Sensors*, *23*(6), 3080.

Boyapati, M., & Aygun, R. (2025). Balancergnn: Balancer graph neural networks for imbalanced datasets: A case study on fraud detection. *Neural Networks*, *182*, 106926.

Burkholder, K., Kwock, K., Xu, Y., Liu, J., Chen, C., & Xie, S. (2021). Certification and trade-off of multiple fairness criteria in graph-based spam detection. In *30th acm cikm* (pp. 130–139).

Buyl, M., & De Bie, T. (2020). Debayes: a bayesian method for debiasing network embeddings. In *Icml* (pp. 1220–1229).

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes*, *11*(4), 1210.

Chen, H., He, Y., Ji, J., & Shi, Y. (2019). A machine learning method for identifying critical interactions between gene pairs in Alzheimer's disease prediction. *Frontiers in Neurology*, *10*, 1162.

Chen, X., Zhang, B., Wang, T., Bonni, A., & Zhao, G. (2020). Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. *BMC Bioinformatics*, *21*(1), 269.

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2020). Graph neural network for fraud detection via spatial-temporal attention. *IEEE Transactions on Knowledge and Data Engineering*, *34*, 3800–3813.

Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph neural networks for financial fraud detection: a review. *Frontiers of Computer Science*, *19*(9), 199609.

Cho, K. (2019). Emerging roles of complement protein C1q in neurodegeneration. *Aging and Disease*, *10*(3), 652–663.

Cui, Y., Han, X., Chen, J., Zhang, X., Yang, J., & Zhang, X. (2025). Fraudgnn-rl: a graph neural network with reinforcement learning for adaptive financial fraud detection. *IEEE Open Journal of the Computer Society*.

Dai, E., & Wang, S. (2021). Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In *Acm wsdm* (pp. 680–688).

Data61, C. (2018). *Stellargraph machine learning library.* `https://github.com/stellargraph/stellargraph`. GitHub.

Dong, G., Tang, M., Wang, Z., Gao, J., Guo, S., Cai, L., ... Boukhechba, M. (2023). Graph neural networks in iot: A survey. *ACM TOSN*, *19*(2), 1–50.

Dong, Y., Kang, J., Tong, H., & Li, J. (2021). Individual fairness for graph neural networks: A ranking based approach. In *27th acm sigkdd* (pp. 300–310).

Dong, Y., Liu, N., Jalaian, B., & Li, J. (2022). Edits: Modeling and mitigating data bias for graph neural networks. In *the acm web conference* (pp. 1259–1269).

Dong, Y., Ma, J., Wang, S., Chen, C., & Li, J. (2023). Fairness in graph mining: A survey. *TKDE*, 1–22.

Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *29th acm international conference on information & knowledge management* (pp. 315–324).

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), eaao5580.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *3rd itcs* (pp. 214–226).

Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., & Yin, D. (2019). Graph neural networks for social recommendation. In *Www* (pp. 417–426).

Feng, Z., Wang, R., Wang, T., Song, M., Wu, S., & He, S. (2025). A comprehensive survey of dynamic graph neural networks: Models, frameworks, benchmarks, experiments and challenges. *IEEE Transactions on Knowledge and Data Engineering*.

Fu, K., Cheng, D., Tu, Y., & Zhang, L. (2016). Credit card fraud detection using convolutional neural networks. In *International conference on neural information processing* (pp. 483–490).

Galvan, M. D., Greenlee-Wacker, M. C., & Bohlson, S. S. (2012). C1q and phagocytosis: the perfect complement to a good meal. *Journal of Leukocyte Biology*, *92*(3), 489–497.

Gao, C., Yin, S., Wang, H., Wang, Z., Du, Z., & Li, X. (2023). Medical-knowledge-based graph neural network for medication combination prediction. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 13246–13257.

Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., ... others (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*, *1*(1), 1–51.

Gao, Y., Zhang, X., Sun, Z., Chandak, P., Bu, J., & Wang, H. (2025). Precision adverse drug reactions prediction with heterogeneous graph neural network. *Advanced Science*, *12*(4), 2404671.

Ge, D., Gu, J., Chang, S., & Cai, J. (2020). Credit card fraud detection using lightgbm model. In *International conference on e-commerce and internet technology (ecit)* (pp. 232–236).

Gillies, M., Arnold, J., Bhandari, S., Essex, R. W., Young, S., Squirrell, D., ... Barthelmes, D. (2020). Ten-year treatment outcomes of neovascular age-related

macular degeneration from two regions. *American Journal of Ophthalmology*, *210*, 116–124.

Gu, Y., Peng, S., Li, Y., Gao, L., & Dong, Y. (2025). Fc-hgnn: A heterogeneous graph neural network based on brain functional connectivity for mental disorder identification. *Information Fusion*, *113*, 102619.

Gupta, R., Kleinjans, J., & Caiment, F. (2021). Identifying novel transcript biomarkers for hepatocellular carcinoma (HCC) using RNA-Seq datasets and machine learning. *BMC Cancer*, *21*(1), 962.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, *30*.

Han, L., Wang, L., Cheng, Z., Wang, B., Yang, G., Cheng, D., & Lin, X. (2025). Mitigating the tail effect in fraud detection by community enhanced multi-relation graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in NIPS*, *29*.

Hasegawa, E., Sweigard, H., Husain, D., Olivares, A. M., Chang, B., Smith, K. E., ... Connor, K. M. (2014). Characterization of a spontaneous retinal neovascular mouse model. *PLoS One*, *9*(9), e106507.

Hofmann, H. (1994). *Statlog (German Credit Data).* UCI Machine Learning Repository. (https://doi.org/10.24432/C5NC77)

Husejinovic, A. (2020). Credit card fraud detection using naive bayesian and c4. 5 decision tree classifiers. *Periodicals of Engineering and Natural Sciences*, *8*(1), 1–5.

Hussein, A. S., Khairy, R. S., Najeeb, S. M. M., & Alrikabi, H. T. S. (2021). Credit card fraud detection using fuzzy rough nearest neighbor and sequential minimal

optimization with logistic regression. *International Journal of Interactive Mobile Technologies*, *15*(5), 24–42.

Innan, N., Sawaika, A., Dhor, A., Dutta, S., Thota, S., Gokal, H., ... Bennai, M. (2024). Financial fraud detection using quantum graph neural networks. *Quantum Machine Intelligence*, *6*(1), 7.

Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, *14*, 100189.

Itoo, F., Singh, S., et al. (2021). Comparison and analysis of logistic regression, naïve bayes and knn machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, *13*(4), 1503–1511.

Jemima Jebaseeli, T., Venkatesan, R., & Ramalakshmi, K. (2021). Fraud detection for credit card transactions using random forest algorithm. In *Intelligence in big data technologies—beyond the hype* (pp. 189–197). Springer.

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514.

Jiang, W., & Luo, J. (2022). Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, *207*, 117921.

Jin, J., Zhu, T., & Li, C. (2025). Graph neural network-based prediction framework for protein-ligand binding affinity: A case study on pediatric gastrointestinal disease targets. *Journal of Medicine and Life Sciences*, *1*(3), 136–142.

Jing, R., Tian, H., Zhou, G., Zhang, X., Zheng, X., & Zeng, D. D. (2021). A gnn-based few-shot learning model on the credit card fraud detection. In *1st international conference on digital twins and parallel intelligence (dtpi)* (pp. 320–323).

Ju, W., Yi, S., Wang, Y., Xiao, Z., Mao, Z., Li, H., ... others (2025). A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234–245.

Kang, J., He, J., Maciejewski, R., & Tong, H. (2020). Inform: Individual fairness on graph mining. In *26th acm sigkdd* (pp. 379–389).

Kauppinen, A., Paterno, J. J., Blasiak, J., Salminen, A., & Kaarniranta, K. (2016). Inflammation and its role in age-related macular degeneration. *Cellular and Molecular Life Sciences*, *73*, 1765–1786.

Khachigian, L. M., Liew, G., Teo, K. Y., Wong, T. Y., & Mitchell, P. (2023). Emerging therapeutic strategies for unmet need in neovascular age-related macular degeneration. *Journal of Translational Medicine*, *21*(1), 133.

Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, *2023*(1), 1406060.

Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, *11*(1), 18.

Kumar, M. S., Soundarya, V., Kavitha, S., Keerthika, E., & Aswini, E. (2019). Credit card fraud detection using random forest algorithm. In *3rd international conference on computing and communications technologies (iccct)* (pp. 149–153).

Lahoti, P., Gummadi, K. P., & Weikum, G. (2019). Operationalizing individual fairness with pairwise fair representations. *preprint arXiv:1907.01439*.

Lee, S., Jung, J., Park, I., Park, K., & Kim, D.-S. (2020). A deep learning and similarity-based hierarchical clustering approach for pathological stage prediction of papillary renal cell carcinoma. *Computational and Structural Biotechnology Journal*, *18*, 2639–2646.

Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., & Ntoutsi, E. (2022). A survey on datasets for fairness-aware machine learning. *DMKD*, *12*(3), 1452–1511.

Li, C.-T., Tsai, Y.-C., Chen, C.-Y., & Liao, J. C. (2025). Graph neural networks for tabular data learning: A survey with taxonomy and directions. *ACM Computing Surveys*, *58*(1), 1–51.

Li, R., Yuan, X., Radfar, M., Marendy, P., Ni, W., O'Brien, T. J., & Casillas-Espinosa, P. M. (2021). Graph signal processing, graph neural network and graph learning on biological data: a systematic review. *IEEE Reviews in Biomedical Engineering*, *16*, 109–135.

Li, Y., Zhang, G., Wang, P., Yu, Z.-G., & Huang, G. (2022). Graph neural networks in biomedical data: a review. *Current Bioinformatics*, *17*(6), 483–492.

Linder, M., Bennink, L., Foxton, R. H., Kirkness, M., & Westenskow, P. D. (2024). In vivo monitoring of active subretinal fibrosis in mice using collagen hybridizing peptides. *Lab Animal*, *53*(8), 196–204.

Liu, G., Tang, J., Tian, Y., & Wang, J. (2021). Graph neural network for credit card fraud detection. In *International conference on cyber-physical social intelligence (iccsi)* (pp. 1–6).

Liu, X., Yan, K., Kara, L. B., & Nie, Z. (2021). Ccfd-net: A novel deep learning model for credit card fraud detection. In *22nd international conference on information reuse and integration for data science (iri)* (pp. 9–16).

Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H., & He, Q. (2021). Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In *the web conference* (pp. 3168–3177).

Loaeza-Reyes, K. J., Zenteno, E., Moreno-Rodríguez, A., Torres-Rosas, R., Argueta-Figueroa, L., Salinas-Marín, R., ... Cervera, Y. P. (2021). An overview of glycosylation and its impact on cardiovascular health and disease. *Frontiers in Molecular Biosciences*, *8*, 751637.

Lu, M., Han, Z., Zhang, Z., Zhao, Y., & Shan, Y. (2021). Graph neural networks in real-time fraud detection with lambda architecture. *preprint arXiv:2110.04559*.

Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., & Calabretto, S. (2020). Towards automated feature engineering for credit card fraud detection using multi-perspective hmms. *Future Generation Computer Systems*, *102*, 393–402.

Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231.

Ma, J., Deng, J., & Mei, Q. (2021). Subgroup generalization and fairness of graph neural networks. *Advances in NIPS*, *34*, 1048–1061.

Ma, Y., Ding, X., Shao, M., Qiu, Y., Li, S., Cao, W., & Xu, G. (2022). Association of serum complement C1q and C3 level with age-related macular degeneration in women. *Journal of Inflammation Research*, 285–294.

Ma, Z., Huang, Z., Zhang, C., Liu, X., Zhang, J., Shu, H., ... others (2023). Hepatic Acat2 overexpression promotes systemic cholesterol metabolism and adipose lipid metabolism in mice. *Diabetologia*, *66*(2), 390–405.

Mallone, F., Costi, R., Marenco, M., Plateroti, R., Minni, A., Attanasio, G., ... Lambiase, A. (2021). Understanding drivers of ocular fibrosis: current and future

therapeutic perspectives. *International Journal of Molecular Sciences*, *22*(21), 11748.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *CSUR*, *54*(6), 1–35.

Meng, C., Zhou, L., & Liu, B. (2020). A case study in credit fraud detection with smote and xgboost. In *Journal of physics: Conference series* (Vol. 1601, p. 052016).

Mi, X., Zou, B., Zou, F., & Hu, J. (2021). Permutation-based identification of important biomarkers for complex diseases via machine learning models. *Nature Communications*, *12*(1), 3008.

Mohi ud din, A., & Qureshi, S. (2023). A review of challenges and solutions in the design and implementation of deep graph neural networks. *International Journal of Computers and Applications*, *45*(3), 221–230.

Moser, T., Kuehberger, S., Lazzeri, I., Vlachos, G., & Heitzer, E. (2023). Bridging biological cfdna features and machine learning approaches. *Trends in Genetics*, *39*(4), 285–307.

Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, *240*, 122156.

Munikoti, S., Agarwal, D., Das, L., Halappanavar, M., & Natarajan, B. (2023). Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications. *IEEE transactions on neural networks and learning systems*, *35*(11), 15051–15071.

Nagai, N., von Leithner, P. L., Izumi-Nagai, K., Hosking, B., Chang, B., Hurd, R., ... Shima, D. T. (2014). Spontaneous CNV in a novel mutant mouse is associated with Early VEGF-A-driven angiogenesis and late-stage focal edema,

neural cell loss, and dysfunction. *Investigative Ophthalmology & Visual Science*, *55*(6), 3709–3719.

Nagare, M., Ayachit, M., Agnihotri, A., Schwab, W., & Joshi, R. (2021). Glycosyltransferases: the multifaceted enzymatic regulator in insects. *Insect Molecular Biology*, *30*(2), 123–137.

Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020). Credit card fraud detection based on machine and deep learning. In *11th international conference on information and communication systems (icics)* (pp. 204–208).

Nandan, M., Mitra, S., & De, D. (2025). Graphxai: a survey of graph neural networks (gnns) for explainable ai (xai). *Neural Computing and Applications*, 1–52.

Network, A. P. (2022, December). *Payment fraud statistics Jul 2021-Jun 2022.* Retrieved from `https://www.auspaynet.com.au/resources/fraud-statistics/July-2021-June-2022`

Nita, M., Strzałka-Mrozik, B., Grzybowski, A., Mazurek, U., & Romaniuk, W. (2014). Age-related macular degeneration and changes in the extracellular matrix. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *20*, 1003.

Paul, S. G., Saha, A., Hasan, M. Z., Noori, S. R. H., & Moustafa, A. (2024). A systematic review of graph neural network in healthcare-based applications: Recent advances, trends, and future directions. *IEEE Access*, *12*, 15145–15170.

Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *CSUR*, *55*(3), 1–44.

Pfeifer, B., Saranti, A., & Holzinger, A. (2022). Gnn-subnet: disease subnetwork detection with explainable graph neural networks. *Bioinformatics*, *38*(Supplement_2), ii120–ii126.

Pham, P., Bui, Q.-T., Nguyen, N. T., Kozma, R., Yu, P. S., & Vo, B. (2025). Topological data analysis in graph neural networks: Surveys and perspectives. *IEEE Transactions on Neural Networks and Learning Systems*.

Pun, F. W., Ozerov, I. V., & Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*, *44*(9), 561–572.

Rafique, R., Islam, S. R., & Kazi, J. U. (2021). Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*, *19*, 4003–4017.

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using adaboost and majority voting. *IEEE Access*, *6*, 14277–14284.

Rao, Y., Mi, X., Duan, C., Ren, X., Cheng, J., Chen, Y., ... Wei, X. (2021). Know-Gnn: An explainable knowledge-guided graph neural network for fraud detection. In *International conference on neural information processing* (pp. 159–167).

Report, T. N. (2021, December). *Card fraud worldwide - the nilson report.* Retrieved from `https://nilsonreport.com/mention/1515/1link/`

Rossato, F. A., Su, Y., Mackey, A., & Ng, Y. S. E. (2020). Fibrotic changes and endothelial-to-mesenchymal transition promoted by VEGFR2 antagonism alter the therapeutic effects of VEGFA pathway blockage in a mouse model of choroidal neovascularization. *Cells*, *9*(9), 2057.

Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018). Deep learning detecting fraud in credit card transactions. In *Systems and information engineering design symposium (sieds)* (pp. 129–134).

Sahin, Y., & Duman, E. (2010). Detecting credit card fraud by decision trees and support vector machines. In *World congress on engineering* (Vol. 2188, pp. 442–447).

Sarkar, S., & Alvari, H. (2020). Mitigating bias in online microfinance platforms: A case study on kiva. org. In *Ecml pkdd* (pp. 75–91).

Schlichtkrull, M., Kipf, T. N., Bloem, P., Berg, R. v. d., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In *European semantic web conference* (pp. 593–607).

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, *10*(07), 3897–3904.

Sharma, K., Lee, Y.-C., Nambi, S., Salian, A., Shah, S., Kim, S.-W., & Kumar, S. (2024). A survey of graph neural networks for social recommender systems. *ACM Computing Surveys*, *56*(10), 1–34.

Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*, 105458.

Shughoury, A., Sevgi, D. D., & Ciulla, T. A. (2022). Molecular genetic mechanisms in age-related macular degeneration. *Genes*, *13*(7), 1233.

Singh, A., & Jain, A. (2019). Adaptive credit card fraud detection techniques based on feature selection method. In *Advances in computer communication and computational sciences* (pp. 167–178). Springer.

Slovin, S., Carissimo, A., Panariello, F., Grimaldi, A., Bouché, V., Gambardella, G., & Cacchiarelli, D. (2021). Single-cell RNA sequencing analysis: a step-by-step overview. *RNA Bioinformatics*, 343–365.

Song, W., Dong, Y., Liu, N., & Li, J. (2022). Guide: Group equality informed individual fairness in graph neural networks. In *28th acm sigkdd* (pp. 1625–1634).

Song, Z., Yang, X., Xu, Z., & King, I. (2022). Graph-based semi-supervised learning: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(11), 8174–8194.

Strokach, A., Becerra, D., Corbi-Verge, C., Perez-Riba, A., & Kim, P. M. (2020). Fast and flexible protein design using deep graph neural networks. *Cell Systems*, *11*(4), 402–411.

Sun, F., Sun, J., & Zhao, Q. (2022). A deep learning method for predicting metabolite–disease associations via graph neural network. *Briefings in bioinformatics*, *23*(4), bbac266.

Tenbrock, L., Wolf, J., Boneva, S., Schlecht, A., Agostini, H., Wieghofer, P., ... Lange, C. (2022). Subretinal fibrosis in neovascular age-related macular degeneration: current concepts, therapeutic avenues, and future perspectives. *Cell and Tissue Research*, *387*(3), 361–375.

Veličković, P. (2023). Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, *79*, 102538.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al. (2017). Graph attention networks. *Stat*, *1050*(20), 10–48550.

Venkat, V., Abdelhalim, H., DeGroat, W., Zeeshan, S., & Ahmed, Z. (2023). Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine. *Genomics*, *115*(2), 110584.

Wang, D., Lin, J., Cui, P., Jia, Q., Wang, Z., Fang, Y., ... Qi, Y. (2019). A semi-supervised graph attentive network for financial fraud detection. In *Icdm* (pp. 598–607).

Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal heterogeneous graph contrastive learning for fraud detection in credit card transactions. *IEEE Access*.

Wang, J., Zhang, S., Xiao, Y., & Song, R. (2021). A review on graph neural network methods in financial applications. *preprint arXiv:2111.15367*.

Wang, Y., Kang, J., Xia, Y., Luo, J., & Tong, H. (2022). ifig: Individually fair multi-view graph clustering. In *Bigdata* (pp. 329–338).

Wang, Y., Wu, M., Li, X., Xie, L., & Chen, Z. (2025). A survey on graph neural networks for remaining useful life prediction: Methodologies, evaluation and future trends. *Mechanical Systems and Signal Processing*, *229*, 112449.

Welling, M., & Kipf, T. N. (2017). Semi-supervised classification with graph convolutional networks. In *International conference on learning representations (iclr)*.

Won, J., Shi, L. Y., Hicks, W., Wang, J., Hurd, R., Naggert, J. K., ... Nishina, P. M. (2011). Mouse model resources for vision research. *Journal of Ophthalmology*, *2011*(1), 391384.

Wu, L., Cui, P., Pei, J., Zhao, L., & Guo, X. (2022). Graph neural networks: foundation, frontiers and applications. In *Proceedings of the 28th acm sigkdd conference on knowledge discovery and data mining* (pp. 4840–4841).

Wu, Y.-N., Su, X., Wang, X.-Q., Liu, N.-N., & Xu, Z.-W. (2023). The roles of phospholipase C-β related signals in the proliferation, metastasis and angiogenesis of malignant tumors, and the corresponding protective measures. *Frontiers in Oncology*, *13*, 1231875.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, *32*(1), 4–24.

Xiong, J., Xiong, Z., Chen, K., Jiang, H., & Zheng, M. (2021). Graph neural networks for automated de novo drug design. *Drug Discovery Today*, *26*(6), 1382–1393.

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks? *preprint arXiv:1810.00826*.

Yang, L., Wang, S., Tao, Y., Sun, J., Liu, X., Yu, P. S., & Wang, T. (2023). Dgrec: Graph neural network for recommendation with diversified embedding generation. In *Proceedings of the sixteenth acm international conference on web search and data mining* (pp. 661–669).

Yao, L., Mao, C., & Luo, Y. (2019). Graph convolutional networks for text classification. In *Aaai* (Vol. 33, pp. 7370–7377).

Yeh, I.-C., & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473–2480.

Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., & Leskovec, J. (2018). Graph convolutional neural networks for web-scale recommender systems. In *24th acm sigkdd* (pp. 974–983).

Yu, Z., Wang, Z., Yu, X., & Zhang, Z. (2020). RNA-Seq-Based Breast Cancer Subtypes Classification Using Machine Learning Approaches. *Computational Intelligence and Neuroscience*, *2020*(1), 4737969.

Zakaria, R. M., Rahman, M. M., Rahman, H., Rafi, M. A., et al. (2025). Detecting financial fraud in real-time transactions using graph neural networks and anomaly detection techniques. *Journal of Economics, Finance and Accounting Studies*, *7*(6), 01–13.

Zhang, G., Li, Z., Huang, J., Wu, J., Zhou, C., Yang, J., & Gao, J. (2022). efraudcom: An e-commerce fraud detection system via competitive graph neural networks. *ACM Transactions on Information Systems (TOIS)*, *40*(3), 1–29.

Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, *557*, 302–316.

Zhang, X., Jonassen, I., & Goksøyr, A. (2021). Machine learning approaches for biomarker discovery using gene expression data. *Bioinformatics*.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, *1*, 57–81.

Zhu, L., Jones, C., & Zhang, G. (2018). The role of phospholipase C signaling in macrophage-mediated inflammatory response. *Journal of Immunology Research*, *2018*(1), 5201759.