# Advanced Deep Learning Techniques for EEG based Sleep Signal Analysis

WEI QU

Doctor of Philosophy

Supervisor: Professor Zhiyong Wang
Associate Supervisor: Professor Christopher Gordon

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

School of Computer Science
Faculty of Engineering
The University of Sydney
Australia

25 February 2025

## Statement 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 purpose.

The intellectual content presented in this thesis is the result of my individual effort and I have acknowledged all aid and resources used in its preparation.

Student Name: Wei Qu
Signature,
Wei Qu
Date: 25 February 2025

## Authorship Attribution Statement

This thesis is mainly based on the following works:

1. **Wei Qu**, Zhiyong Wang, Hong Hong, Zheru Chi, David Dagan Feng, Ron Grunstein, and Christopher Gordon. (2020). "A Residual based Attention Model for EEG based Sleep Staging." *IEEE Journal of Biomedical and Health Informatics*, 24(12), 2833-2843.

Chapter 3 of this thesis is based on this work. I (Wei Qu) assumed primary responsibility for method development and implementation, data processing, experiments, algorithm evaluation, and drafting and revising the manuscript under the supervision of Professor Zhiyong Wang and Professor Christopher Gordon. Other co-authors contributed to method conceptualization, experimental design, experiment analysis, and manuscript revision.

2. **Wei Qu**, Chien-Hui Kao, Hong Hong, Zheru Chi, Ron Grunstein, Christopher Gordon, and Zhiyong Wang (2021). "Single-channel EEG based insomnia detection with domain adaptation." *Computers in Biology and Medicine*, 139, 104989.

Chapter 4 of this thesis is derived from this work. My contributions included conducting the data investigation, performing an extended comparison study, and leading the formulation and compilation of the final data analysis under the supervision of Professor Zhiyong Wang and Professor Christopher Gordon. Other co-authors contributed to method conceptulisation, experimental design, experiment analysis, and manuscript revision. Dr. Chien-Hui Kao also assisted in data collection.

3. **Wei Qu**, Wenxi Yue, Jeremiah Deng, Christopher Gordon, Ron Grunstein and Zhiyong Wang (2025). "CAPSleepNet: A Cyclic Alternating Pattern based Multi-view Sleep Staging Model." (In preparation)

Chapter 5 of this thesis is based on this work. I spearheaded the development and implementation of the benchmarking framework, drafting and revising the manuscript, under

the supervision of Professor Zhiyong Wang and Professor Christopher Gordon. Other co-authors contributed to method conceptualization, experimental design, experiment analysis, and manuscript revision. Dr. Wenxi Yue also helped the design of all the figures.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Student Name: Wei Qu
Signature,
Wei Qu
Date: 25 February 2025

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

Supervisor Name: Zhiyong Wang
Signature,
Zhiyong Wang
Date: 25 February 2025

# Acknowledgements

This dissertation marks the pinnacle of my most significant scholarly pursuit, a path that has fostered both intellectual development and meaningful connections. The completion of this research was made possible through the support of many individuals who deserve recognition.

I express my deepest gratitude to my supervisor, Prof. Zhiyong Wang, and co-supervisor, Prof. Christopher Gordon, whose steadfast guidance has been instrumental throughout this process. Their mentorship has been transformative in shaping my academic trajectory, challenging conventional boundaries, and nurturing scholarly curiosity. The countless discussions, from dawn meetings with Prof. Wang to evening dialogues with Prof. Christopher, provided consistent encouragement and support over the years.

I extend sincere appreciation to my peers and associates at the Multimedia Computing Laboratory, along with research collaborators, including A/Prof. Hong Hong, A/Prof. Zheru Chi, Prof. David Feng, and Prof. Ron Grunstein for their constructive input and scholarly fellowship. Special acknowledgment goes to fellow researchers Wenxi Yue, Kun Hu, Chien-Hui Kao, and Renfei Sun, whose friendship enhanced my university experience. Their shared moments of academic discourse, personal conversations, and communal laughter brought vitality and motivation to my journey.

My research benefited from the generous financial support of the CRC for Alertness, Safety and Productivity, which was essential for its completion.

I acknowledge an immeasurable debt to my family for their foundational support. Their endless love and encouragement helped me overcome every obstacle. I dedicate this work to my wife, Na He, whose constant support and numerous sacrifices have provided stability throughout my academic pursuit. Finally, I reserve my deepest thanks to my children, Austin and Betty, whose presence brought perpetual light to this journey. Their joy and laughter

served as continuous reminders of life's true priorities, infusing these years with meaning and delight.

# Abstract

Sleep signals are crucial in understanding human health, particularly in diagnosing sleep disorders and analyzing sleep patterns. Polysomnography is the gold standard for measuring objective sleep using a variety of physiological measurements, with the electroencephogram (EEG) determining brain wave activity. The EEG signal along with other measures, such as breathing, heart rate and eye and muscle movement are used to classify sleep staging and detection of sleep disorders. Sleep studies are used extensively in research and clinical practice but typically rely on manual sleep staging, which require expert domain knowledge, are time consuming and costly.

Recently, owing to the great success of machine learning, various studies have been conducted to automate sleep analysis using these techniques. However, we still see that the gaps are threefold. First, although deep learning based models can produce better results than traditional signal process based approaches, no research combined the traditional signal processing theory in deep learning to improve the model performance. Second, due to the scarcity of public dataset in sleep disorders, an approach to transfer learned sleep features from a large dataset to a small dataset can potentially improve the sleep disorder detection. Lastly, while current sleep staging tasks are based on sleep stage annotations which represent sleep macrostructure, so far no study has been conducted to explore the sleep microstructure in sleep medicine.

To fill the above mentioned gaps, in this thesis we first review the sleep signals that are used in contemporary research and justify that single channel EEG signals should be the focus of our research, as they are less intrusive in collection and are widely used in sleep related research. Then we adopt the Hilbert transform from the signal process theory into the deep learning model architecture to provide a direction in the future sleep staging model design process. After that, we reuse this model architecture in a better characterisation of including insomnia and extract the sleep related features from a large dataset using a sleep

staging task and transfer them to a smaller insomnia dataset using domain adaptation. The work shows a promising approach to future research related to sleep disorders. To explore the possibility of using sleep microstructure in sleep related tasks, we integrate cyclic alternating patterns (CAP), which is a biomarker of sleep microstructure, with the sleep stages to capture finer-grained dynamics from an epoch to improve the model performance.

This thesis contributes to the evolution of automated sleep staging and disorder detection by integrating traditional signal processing theory and sleep microstructure. These advances have significant implications for future sleep medicine, enabling scalable and precise tools for personalized healthcare.

**Keywords**: Sleep signals, EEG, Sleep staging, Insomnia detection, Deep learning, Domain adaptation, Cyclic Alternating Pattern (CAP), Temporal modeling, Attention mechanisms, Sleep microstructure

# List of Acronyms

The following tables provide a list of acronyms used in this thesis to enhance clarity and avoid ambiguity.

TABLE 0.1: List of Acronyms (A-G)

| Acronym | Meaning |
|---|---|
| AASM | American Academy of Sleep Medicine |
| ADC | Analog-to-Digital Converter |
| ADL | Activities of Daily Living |
| AI | Artificial Intelligence |
| ANOVA | Analysis of Variance |
| API | Application Programming Interface |
| AR | Autoregressive Model |
| ARIMA | Autoregressive Integrated Moving Average |
| ASD | Autism Spectrum Disorder |
| ASR | Automatic Speech Recognition |
| BCI | Brain-Computer Interface |
| BVP | Blood Volume Pulse |
| CAP | Cyclic Alternating Pattern |
| CBC | Convolutional Block Channel |
| CDN | Content Delivery Network |
| CF | Cross-Frequency Coupling |
| CNC | Computerized Numerical Control |
| CNN | Convolutional Neural Network |
| COPD | Chronic Obstructive Pulmonary Disease |
| CRF | Conditional Random Field |
| CT | Computed Tomography |
| CTH | Continuous Theta Burst Stimulation |
| CV | Cross-Validation |

TABLE 0.2: List of Acronyms (H-N)

| **Acronym** | **Meaning** |
|---|---|
| DA | Domain Adaptation |
| DB | Database |
| DCNN | Deep Convolutional Neural Network |
| DIDA | Dynamic Instance Domain Adaptation |
| DS | Dataset |
| DTW | Dynamic Time Warping |
| ECG | Electrocardiogram |
| EEG | Electroencephalogram |
| EMG | Electromyogram |
| ENA | Environmental Neurotechnology Applications |
| EOG | Electrooculogram |
| ER | Emergency Room |
| ESD | European Sleep Database |
| F1 | F1 Score |
| FFT | Fast Fourier Transform |
| FDA | Food and Drug Administration |
| FIR | Finite Impulse Response |
| FMRI | Functional Magnetic Resonance Imaging |
| FT | Feature Transformation |
| GC | Granger Causality |
| GMM | Gaussian Mixture Model |
| GPU | Graphics Processing Unit |

TABLE 0.3: List of Acronyms (O-Z)

| Acronym | Meaning |
|---|---|
| GRU | Gated Recurrent Unit |
| HBM | Hierarchical Bayesian Model |
| HCP | Human Connectome Project |
| HMM | Hidden Markov Model |
| HR | Heart Rate |
| HRV | Heart Rate Variability |
| ICA | Independent Component Analysis |
| ICD | International Classification of Diseases |
| ICU | Intensive Care Unit |
| IDF | Inverse Document Frequency |
| ILC | Intraoperative Language Center |
| IMF | Intrinsic Mode Function |
| IoT | Internet of Things |
| IPR | Instantaneous Phase Resetting |
| IRB | Institutional Review Board |
| IRDA | International Repository of Diabetes Data |
| ISDE | International Society for Disease Surveillance |
| K-NN | K-Nearest Neighbors |
| LBP | Local Binary Pattern |
| LDA | Linear Discriminant Analysis |
| LDS | Linear Dynamical Systems |
| LFP | Local Field Potential |
| LSTM | Long Short-Term Memory |
| LVM | Latent Variable Model |
| MFCC | Mel-Frequency Cepstral Coefficients |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| MRI | Magnetic Resonance Imaging |
| MSDA | Multi-Source Domain Adaptation |
| MSU | Michigan State University |
| NREM | Non-Rapid Eye Movement |
| SOTA | State Of The Art |

# Contents

CHAPTER 1

# Introduction

---

Sleep is essential to human health, and disorders in sleep, such as insomnia, can deteriorate an individual's health condition. Research has found that disruption of normal sleep patterns could be very helpful in the diagnosis and administration of medications in neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and multiple sclerosis [1, 2, 3]. Therefore, collecting sleep signals during sleep monitoring and analyzing them becomes increasingly important in sleep studies.

Traditionally, overnight polysomnography (PSG) recordings are the foundation of most sleep studies [4]. It comprises electro-oculogram (EOG), electromyography (EMG), electroencephalogram (EEG), electrocardiogram (ECG), oxygen saturation, airflow, and signals from other sensors. Each of those signals monitors activities from various parts of the body, such as the heart and brain through the electrodes contacted to the patient's scalp or skin. Those signals can have different uses, such as EEG could be used for sleep staging, which is a task to classify 30 seconds EEG segments into different sleep stages while ECG could help detect sleep disorders such as obstructive sleep apnea (OSA) which is reflected as difficulties in breathing during sleep. However, these tasks must be performed by highly trained sleep technicians. To efficiently utilize those signals, several studies have recently been carried out to develop automatic diagnostic tools with sleep signals.

On the other hand, some sleep diseases, such as insomnia, have numerous subtypes, and identifying which subtype the patient has is challenging. With the rapidly increasing availability of electronic health records (EHRs), identifying phenotypes of a specific disease becomes possible, but typical phenotyping approaches require intensive intervention from medical experts. Although researchers have proposed many algorithms and models to help

phenotyping on top of EHR data, extracting explainable phenotypes is still a challenging problem to solve.

This dissertation focuses on sleep staging and phenotyping in insomnia. The research involves investigating the robustness of integrating various datasets and using deep learning models to improve the learning of sleep-related representations that can be used to improve the accuracy of classification.

## 1.1 Research Motivations

This research aims to investigate the development of algorithms that can use deep learning techniques in single-channel EEG signals to classify sleep stages and detect insomnia. The three primary questions driving this research are as follows:

- How can signal processing theory be integrated into deep learning architectures to create explainable models that improve sleep staging classification performance?
- How can we adapt the features learned from the popular sleep staging dataset to other datasets to make an automatic diagnosis of sleep disorders, e.g. insomnia, possible? If the performance of the model is good, can we find some interesting features that contribute to the phenotyping of sleep disorders?
- Although sleep stages are representations of sleep macrostructure, can we leverage the sleep microstructure into the sleep staging model as well to improve performance?

This thesis addresses the classification of sleep stages and the detection of insomnia from single-channel EEG signals using deep learning techniques. It is centered on three main research problems mentioned previously:

**The first research problem:** Deep learning approaches are becoming popular due to their performance. However, it is less explainable compared to the signal processing approach. The classification models for sleep staging in existing studies depend mainly on a single EEG channel that is cost-effective and suitable for real-time monitoring. Although existing research is based either on signal processing or deep learning, this research explores the potential

of incorporating signal processing theories into deep learning to address this problem. The model is expected to provide an explainable architecture design and improve the accuracy of the sleep stage classification.

**The second research problem:** Due to the scarcity of the insomnia data set, it is difficult to perform deep learning-based research. In addition, existing approaches to detecting insomnia from EEG signals are based on feature engineering methods involving complex and time-consuming feature extraction methods. To overcome these issues, this dissertation aims to utilize deep learning techniques to automatically extract features from raw data, eliminating the need for laborious manual feature engineering, and to improve the detection of insomnia.

**The third research problem:** Previous studies have limitations in combining different views from signals. These studies focus on composing the frequency domain features, a simple transformation of time domain features that can be obtained easily by deep neural networks. In addition, it does not inject additional information on sleep microstructure in this approach. Furthermore, detecting finer-grained activities within specific sleep stages, such as the non-rapid eye movement (NREM) stage, is not prioritized. Medical studies [5] have shown that the cyclic alternating pattern (CAP) is the marker of sleep microstructure and often provides critical information on sleep quality, especially in NREM sleep. To address the aforementioned issue, this thesis aims to develop a deep learning model that effectively integrates sleep microstructure and macrostructure to accurately detect sleep stages.

## 1.2 Contributions

This thesis presents three distinct solutions to the previously highlighted research problems. Here is a summary of the contributions of this thesis:

- Inspired by the clinical guidelines of sleep staging such as AASM (American Academy of Sleep Medicine) rules where different stages are generally characterized by EEG waveforms of various frequencies, we propose a multi-scale deep architecture by decomposing an EEG signal into different frequency bands as input

to CNNs. To model global temporal context, we utilize the multi-head self-attention module of the transformer model to not only improve performance, but also shorten the training time as well as provide the explainability of the model. In addition, we choose residual based architecture which makes training end-to-end. Experimental results on two widely used sleep staging datasets demonstrate the effectiveness and significant efficiency (up to 12 times less training time) of our proposed method over the state-of-the-art.

- To utilize such abundant EEG datasets to address the data scarcity issue in insomnia detection, in this thesis, we propose a domain adaptation based model to better extract insomnia related features from the target domain by leveraging stage annotations from the source domain. For each domain, two pairs of common encoder and private encoder are first trained to extract sleep related features and sleep irrelevant features, respectively. In order to further discriminate the source domain and the target domain, a domain classifier is introduced. Then, the common encoder of the target domain will be used together with the Long Short Term Memory (LSTM) network for insomnia detection. To the best of our knowledge, this is the first deep learning based domain adaptation model using single channel raw EEG signals to detect insomnia at the subject level. Experimental results on the two target domain datasets demonstrate that our model generalizes well on two target domain datasets with different sampling rates.
- To incorporate the sleep microstructure in sleep staging, we propose a novel method, namely CAPSleepNet, which leverages CAP features as another way to improve performance in sleep staging. Since CAP labels are not readily available in most staging datasets, our method adopts a domain adaptation framework to adapt the epoch features to the CAP features so that they can be transferred from the source CAP dataset to the target staging dataset. Furthermore, to model various discrepancies between the source and target datasets, such as gender, age, sampling rate, etc., we view each sample as a fine-grained domain and turn the domain adaptation into a classification problem. The end-to-end training process is supervised by an inter-epoch Long Short Term Memory (LSTM) stage classifier, which takes both

epoch and CAP features as input and another intra-epoch LSTM CAP classifier simultaneously. We conducted experiments on two public sleep staging datasets, i.e., Montreal Archive of Sleep Studies (MASS) and Sleep-EDF, and demonstrated that our proposed method outperforms the state-of-the-art.

## 1.3 Thesis Structure

The technical materials discussed in the chapters of this thesis have been published or are currently under review. Each chapter focuses on a specific research problem. Chapter 2 provides a systematic review summarizing and analyzing various studies that have utilized machine learning techniques in sleep medicine, particularly those employing EEG. This chapter aims to establish the current state-of-the-art ML-based sleep stage classification, offering a comprehensive overview of methodologies, challenges, and potential advancements in this field.

Chapter 3 directly addresses the first research question by incorporating the concept of Hilbert transform in the deep learning model for sleep stage classification. Examines enhancing accuracy compared to single-channel EEG-based approaches using traditional signal processing theory. This chapter was published in IEEE Journal of Biomedical and Health Informatics, 2020.

Chapter 4 relates to the second research question, which focuses on detecting insomnia, specifically from EEG signals. This chapter extends the previous deep learning architectures. It analyzes how to enrich a small insomnia dataset using the features learned in a large sleep staging dataset to improve the effectiveness of insomnia detection from EEG signals. This chapter was published in Computers in Biology and Medicine in 2021.

Chapter 5 is aligned with the third research question by introducing the CAP as the marker of sleep microstructure, which aims to improve sleep staging accuracy during non-rapid eye movement (NREM) sleep. Investigate how the sleep microstructure improves the detection of

these NREM events compared to existing models. This chapter is under review and has not yet been published.

CHAPTER 2

# Sleep Signal Analytics for Sleep Medicine: A Review

---

Making sleep monitoring less intrusive and expensive to achieve high-quality longitudinal sleep monitoring drives researchers to exploit new methods to collect signals. Wearables with sleep tracking capabilities, such as wrist-worn devices like Fitbit, measure sleep using accelerometry signals to determine sleep and wake. The algorithms in wearable sleep technologies have improved significantly over the last 10 years and research shows that incorporation of multiple physiological signals, such as heart rate, greatly improves the sleep-wake detection accuracy comparable to PSG [6, 7, 8]. Wireless signals from devices such as the Doppler radar [9, 10] and WiFi routers [11] can also carry information related to sleep patterns.

Using signals collected from PSG and other devices, researchers continue to develop automated diagnostic systems to reduce the burden on sleep technicians. For example, a number of models and algorithms are created for the staging of sleep [12, 13, 14, 15] which is categorized EEG segments into different phases of sleep. However, it is usually a headache for new sleep researchers who want to analyze sleep signals but do not know where to start. For example, ECG, which monitors heart, is beneficial in detecting obstructive sleep apnea (OSA), and EMG, which captures actigraphy, is the gold standard in the diagnosis of periodic limb movement (PLM). Surveys can be found for different analytic tasks using sleep signals [16, 17, 18, 19, 20], but no reviews on sleep signals and their applications can be found in our literature search. Hence, providing a systematic overview of the area of sleep signals and combining them with specific analytic tasks becomes the goal of this chapter.

In this chapter, a systematic review is conducted to present recent sleep-related research using signals collected from various devices. We start with an overview of PSG and how it is collected, followed by a section to introduce typical sleep disorders and the diagnosis.

Then the definition of sleep staging will be introduced. After that, research on different analytic tasks using sleep signals is discussed. In addition, to help researchers cross-validate their research with others, an overview of popular public datasets and evaluation metrics is presented. Conclusions and future work are included in the end.

## 2.1 Sleep Signals

Since polysomnography (PSG) and signals derived from PSG are ground truth in most sleep researches [4], how to collect these signals will be discussed in this section. Due to PSG recording being an expensive and intrusive way to monitor sleep which requires massive electrodes to be connected to the patient's head and body, recent alternatives of PSG which are targeting at providing portable, patient-friendly sleep monitoring will be introduced in the rest of the section.

### 2.1.1 PSG signals

PSG comprises electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), electromyography (EMG), oxygen saturation, airflow, and signals from other sensors. Each of them monitors activities from different parts of the body. In this section, an overview of these signals will be given.

#### 2.1.1.1 EEG

EEG captures electrical activities of the brain from multiple electrodes placed on the scalp [21]. Electrodes are named according to the location of placement: frontal (F), temporal (T), central (C), parietal (P), and occipital lobes (O). The electrodes on the lines of the right, left, and middle hemisphere are assigned even numbers, odd numbers, and z, respectively.

Brain waves generated by the EEG system comprise various frequency bands, such as the delta, alpha, theta, and beta wave bands, as shown in Fig. 2.2. Delta waves range from 0.5 to 4 Hz frequency with 20 to 400 $\mu$V amplitude, and occur during low brain activity, such as deep

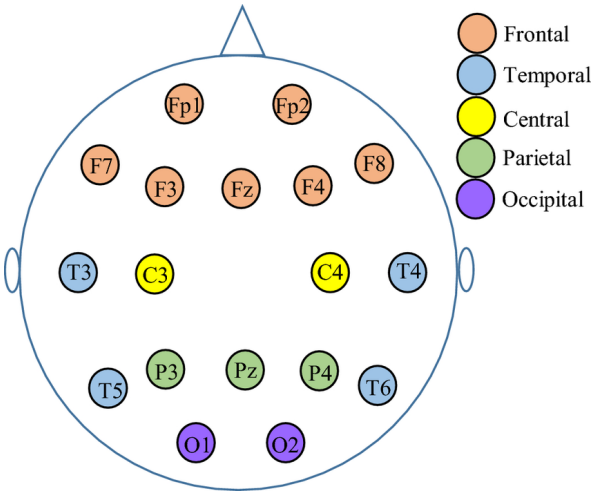


FIGURE 2.1: EEG electrode placement based on traditional 10–20 system (excerpted from [22]).

sleep [23]. Alpha waves range from 8 to 13 Hz frequency with a range of amplitudes of 2 to 10 $\mu$ V. They are usually recorded when a person is awake with closed eyes and in a mental and physical state of rest [24]. Beta waves are recorded at higher frequencies, ranging from 13 to 30 Hz. Their amplitudes range from 1-5 $\mu$V. Beta waves are observed with concentrated attention during the mental working state and also during the REM stage of sleep.

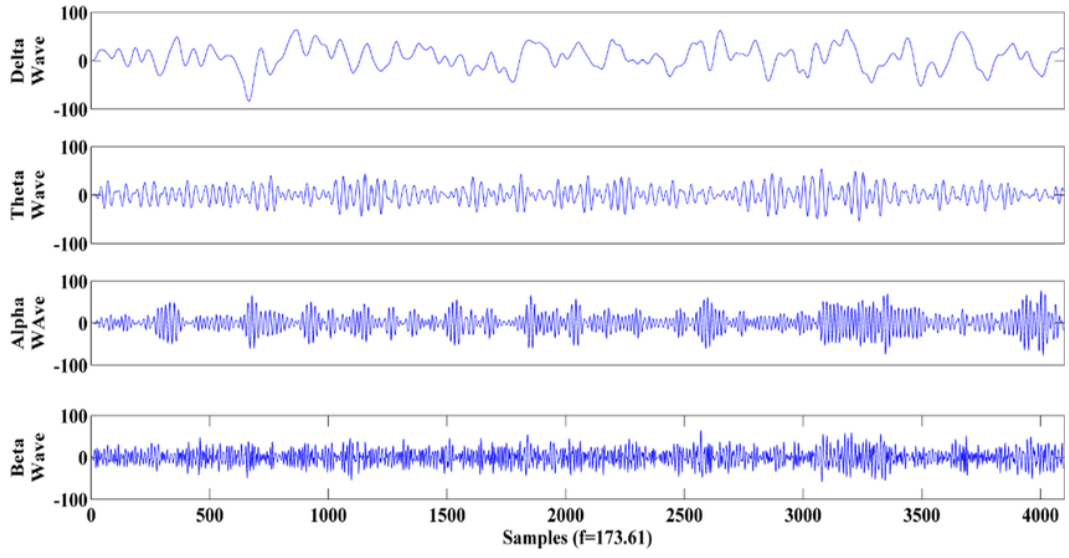


FIGURE 2.2: Typical example of brain waves in EEG (excerpted from [25]).

In some scenarios such as searching for biomarkers, finer-grained EEG signals need to be captured, which requires more electrodes. A high-density EEG electrode placement is shown

below to meet the requirement. It normally provides 256 electrodes with a 1000Hz sampling rate, which can provide finer-grained capture of EEG signals for phenotyping of sleep disease.

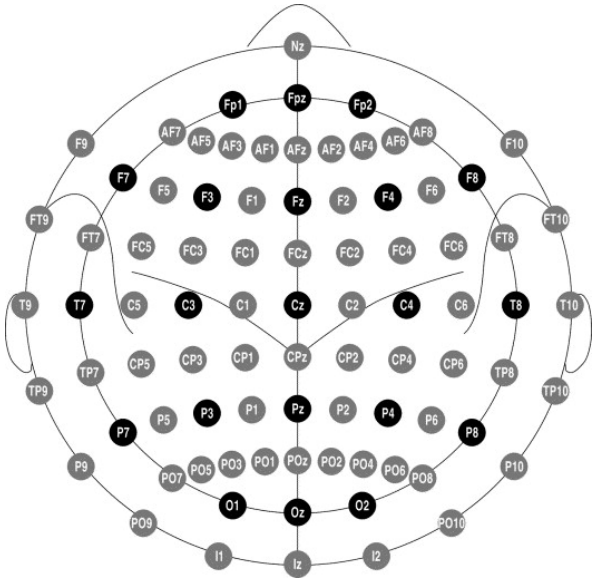


FIGURE 2.3: A sample of high density EEG electrode placement. It compose of 256 electrodes and typically provides higher sampling rate (excerpted from [26]).

#### 2.1.1.2 EOG

EOG which obtains movements of the eyes through electrodes placed around the eyes [27]. In measuring the eye movement, electrode pairs are located either below and above or to the right and left of the eye [27]. An EOG signal response of positive or negative defection can be recorded if the movement of the eye is made from the central position towards one of the electrodes. Therefore, a potential difference would occur when the eye moves between the electrodes placed.

In the AASM standard manual, the positioning of the EOG electrode is recommended approximately 1 cm above and faintly lateral to the outer canthus of one eye (E1 and E2, as shown in Fig. 2.4) and a reference electrode on the mastoid or homolateral ear lobe.

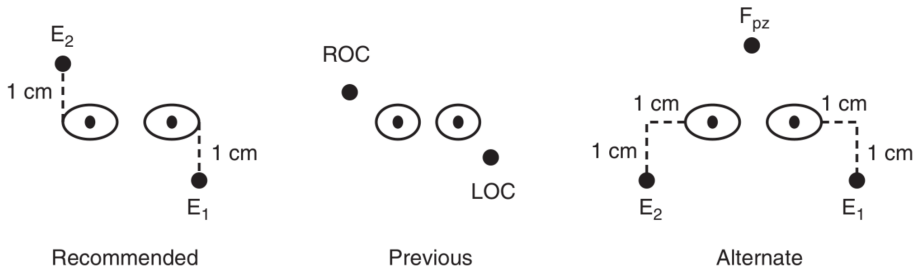


FIGURE 2.4: Recommended, previous and alternate electrode locations of the EOG signal. LOC = left outer canthus; ROC = right outer canthus (excerpted from [28]).

#### 2.1.1.3 EMG

EMG records electrical activities produced by the skeletal muscles [29]. Submental electromyographic (chin) (EMG) is normally used for sleep analysis. Monitoring of chin EMG activity is an essential element only for identifying stage R (REM sleep). The reduction in chin EMG amplitude during REM sleep is a reflection of the generalized skeletal muscle hypotonia present in this stage of sleep.

The placement of EMG electrodes recommended by the AASM scoring manual is illustrated in Fig. 2.5.

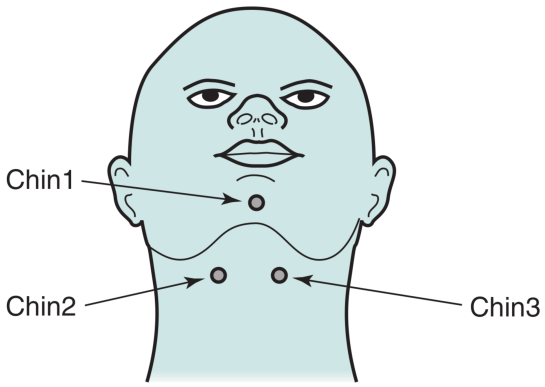


FIGURE 2.5: Submental (chin) EMG electrode positions. Electrode 1 is placed Midline 1 cm above interior edge of mandible. Electrode 2 is placed 2cm below inferior edge of mandible and 2cm right of the midline. The location of electrode 3 is 2cm below inferior edge of mandible and 2 cm left of the midline (excerpted from [28]).

#### 2.1.1.4 ECG

ECG collects electrical activities of the heart via electrodes placed on the skin and can be used to track heartbeat [30]. The ECG is one of the most useful investigations in medicine. Electrodes attached to the chest and/or limbs record small voltage changes as potential difference, which is transposed into a visual tracing. There are 3, 5 and 12 electrode placements that are normally used. A 12 electrode placement is shown below in Fig. 2.6.

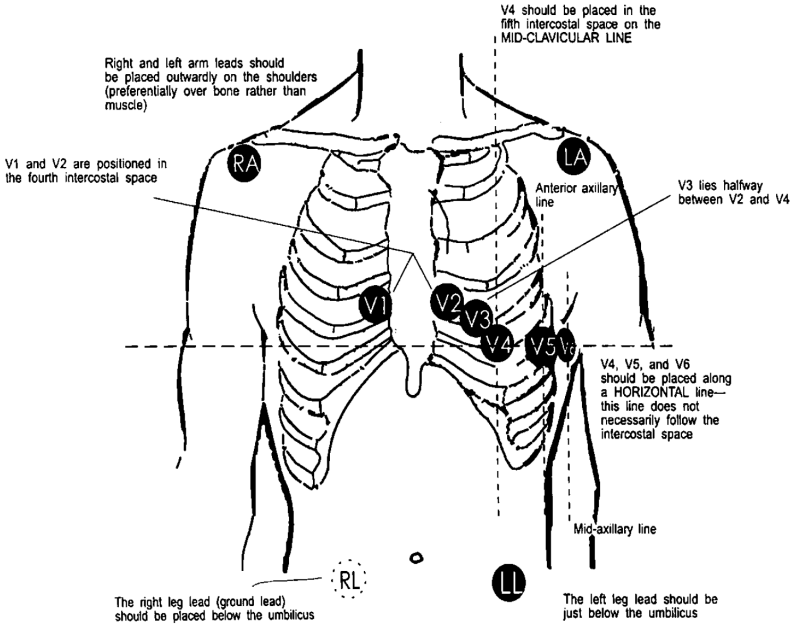


FIGURE 2.6: Placement of 12-lead ECG electrodes. RA indicates right arm; LA, left arm; RL, right leg; and LL, left leg (excerpted from [31]).

### 2.1.2 Wearable and non-contactable signals

PSG recording faces many challenges. It requires well trained technicians, and a large number of electrodes could make patients uncomfortable or intense, while loss of contact and sweat of patients could inject noise and artifacts [32]. Using single-channel EEG or combined with one or more other signals could also achieve good results [33] and alleviate patient discomfort; therefore, researchers began exploring the potential of signals collected from wearable or contactless devices in sleep studies.

Portable alternatives to traditional PSG have been investigated and found the automatic scoring system, showed excellent agreement with PSG manual scoring. Some portable devices could be used to collect cardiac signals, such as heart rate variability (HRV), which is important in the detection of sleep apnea. A wrist photoplethysmography (PPG) device yielded ECG level performance in HRV analysis in [34], while an impedance plethysmography-based wearable sensor device realized high temporal resolution measurements in monitoring movements and cardiac parameters in the wrist in comparison to ECG. In [35], college students were found to tend to sleep late and had a long onset latency according to the average heart rate and breath rate collected from the waistband.

Various sensor technologies enable non-invasive sleep monitoring. Some wearables with built-in accelerometer sensors can record actigraphy which could be used to estimate the estimation of sleep and wake [36] and the time of the sleep period [37]. Research on sleep signals collected from these sensors found that the rural population had relatively better sleep and less sleep disorders than the metropolitan population [38], walking generated better sleep quality and more stable daily blood pressure in seniors [39]. White noise is beneficial in sleep duration according to actigraphy collected from 18 high school students through consumer devices [40]. In addition, individual differences should also be considered in the analysis using actigraphy [41]. Smartphones become more powerful and are equipped with many sensors to interact with humans which could also be utilized in the collection of unobtrusive sleep signals. Acoustic signals from a built-in microphone from a smartphone were used to estimate sleep quality [42] and different statuses of the phone such as locked, shutdown, charging, in darkness, in a stationary state detected from various sensors were used to infer sleep habits in real time [43], which could achieve a comparative result with wearables such as Jawbone wristband and Zeo headband [44]. Combined patterns of smartphone use captured from smartphone sensors with biological signals collected from Fitbit, the authors in [45] found that phone use during bedtime led to significantly worse sleep quality in college students.

Advanced non-contactable monitoring technologies become appealing in modern society. Video is another approach to monitor sleep. The authors in [46] modeled the patient's torso

into two ellipses and captured depth video and audio via the Microsoft Kinect camera for respiratory events classification on top of the dual-ellipses model. A low-cost web camera was used to detect respiratory movements from the luminance difference between pixel levels of a unique bed cover in [47]. Sleep could also be monitored by wireless signals. For example, the WiFi network could track breathing rate and heart rate with fine-grained channels [11] and sleep patterns of patients with REM sleep behavior disorder could be detected from RF signal propagation [48]. Some other novel ways to monitor sleep, such as Doppler radar, which could track heart rate, breathing rate, and physical movement without being unobtrusive, were also useful in estimating the sleep stage [9, 10], while a smart bed system with an integrated electromechanical film-based ballistocardiogram sensor was utilized to quantify restlessness during sleep in [49].

## 2.2 Sleep Staging

Sleep staging is the starting point for most other sleep studies. In this section, an overview of the sleep staging task and challenges it is facing will be given, followed by a discussion of between the traditional and the modern approaches to automate this task.

### 2.2.1 Historical Development of Sleep Stage Classification

The foundational framework for the classification of stages of sleep emerged approximately half a century ago with the seminal work of Rechtschaffen and Kales (R&K) [50]. Their manual introduced systematic rules for defining sleep stages based on a comprehensive analysis of electromyographic (EMG), electroencephalographic (EEG), and electrooculographic (EOG) characteristics. Forty years later, the American Academy of Sleep Medicine (AASM) refined and expanded these original guidelines [51].

According to the AASM's revised classification, sleep can be categorized into three primary stages: Wake (W), Rapid Eye Movement (REM) sleep, and Non-REM (NREM) sleep. NREM sleep is further subdivided into three distinct substages: N1, N2, and N3. Polysomnographic (PSG) recordings serve as the gold standard for assessing the stage of sleep, with specific

neurophysiological markers that define each stage. For example, various neural oscillations, such as alpha, beta, delta, and theta waves, can be segmented from EEG recordings based on their frequency bands [52]. The frequency of occurrence of these waves within 30-second epochs, coupled with EOG measurements, enables precise identification of the sleep stage.

### 2.2.2 Traditional Approaches to Sleep Staging

Traditional sleep staging relies on manual scoring by trained experts, a process with several significant limitations. Manual annotation of 30-second PSG segments is inherently tedious, time-consuming, and subjectively dependent on the interpretation of the individual expert [53]. Despite AASM's standardized guidelines aimed at ensuring interrater consistency [54, 55, 56], interscorer agreement often falls below 90%, fluctuating based on the scorers' experience and fatigue levels [57].

Moreover, PSG recordings present substantial practical challenges. They are costly for patients and logistically complex to implement, particularly in home settings. Consequently, there is an increasing demand for affordable, portable, and unobtrusive automatic sleep staging detection systems that can facilitate high-quality longitudinal sleep monitoring in clinical and home settings.

### 2.2.3 Contemporary Approaches to Sleep Staging

Researchers have pursued multiple methodological approaches to address the challenges of detecting the sleep stage. Signal processing techniques, with their well-established theoretical foundation in frequency domain feature extraction and spectrum analysis, have been a traditional approach to differentiating sleep stages based on neural oscillation characteristics. Currently, deep learning methodologies have emerged as a powerful alternative. Using representation learning capabilities, deep learning algorithms can process raw physiological signals with minimal domain-specific preprocessing, often yielding superior performance. Some research has also explored hybrid approaches that combine signal processing and deep learning techniques to enhance interpretability and accuracy.

### 2.2.4 Signal Processing based Approaches

Research on multisignal integration demonstrates improved sleep stage classification accuracy. Motivated by the benefits of signal combination, Pan et al. [58] developed a Discrete Hidden Markov Model (DHMM) using down-sampled EEG, EOG, and EMG signals processed by eighth-order Butterworth bandpass filters, achieving 85.29% overall agreement in 20 subjects. Helland et al. [15] applied linear discriminant analysis with stepwise selection of characteristics in EEG, ECG, and respiratory signals from ten healthy subjects, achieving precision 93% in epochs annotated by three human experts. Malaekah et al. [23] proposed decision tree-based classifiers to score sleep stages using features in the time and frequency domain extracted from EEG and EOG signals.

Channel reduction strategies address patient comfort and cost considerations. Recognizing the potential to minimize patient discomfort, researchers investigated approaches to reduce signal channels without compromising performance. Stepnowsky et al. [57] compared automated scoring using single-lead EEG with EOG-derived signals from annotated PSG records, finding performance comparable to manual scoring. To optimize feature extraction, Khalighi et al. [12] utilized the maximum overlap wavelet transform (MODWT) to extract time-frequency domain features from preprocessed signals, applying a two-step feature selection algorithm to detect the most discriminative combination of features for EEG, EOG, and EMG channels.

Feature optimization techniques further investigated channel selection methodologies. Piryatinska et al. [59] collected 14 EEG channels from 20 full-term and 16 preterm newborns, exhaustively combining channels to automatically score stages. They ranked automated scoring according to manual scoring to determine the best individual channels. Krakovska et al. [13] employed a forward selection procedure to identify optimal subsets of features from 74 features extracted by bandpass filters, experimenting with entire PSG signals and EEG-alone configurations.

Signal fusion approaches explored advanced integration techniques. Shi et al. [14] proposed a two-stage multiview learning algorithm integrating joint collaborative representation (JCR) and joint sparse representation (JSR) to achieve the fusion of features and classifier levels.

The extracted multiview features were fed into a multikernel learning-based extreme learning machine for sleep stage classification. Liu et al. [60] utilized synchrosqueezing transform (SST) and diffusion map techniques to extract features from two-channel EEG signals, employing a hidden Markov model to predict sleep stages based on features of similar subjects.

Single-channel exploration has emerged as a promising research direction. Boostani et al. [16] provided a comprehensive review of state-of-the-art single channel EEG models, applying five methodologies in sleep data sets to compare performance from various perspectives. They found that the entropy of the wavelet coefficients and the random forest are the most effective combination of features and classifiers. Alickovi et al. [61] applied multiscale principal component analysis to denoise signals and used a discrete wavelet transform to extract informative features. A specific study [62] demonstrated the utility of Jenson-Shannon divergence in discriminating the alpha and theta wave states, efficiently distinguishing between the awake and N1 sleep stages.

### 2.2.5 Deep Learning based Approaches

Temporal context capture drove deep learning methodological innovations. Recognizing the importance of contextual information, Tsinalis et al. [63] developed a convolutional approach that simultaneously processed current epoch signals along with preceding and subsequent epochs. The model used 20 filters with a pooling layer, reshaping filtered signals to capture cross-window information, and addressed data imbalance through class-balanced batch sampling. Sors et al. [64] implemented a 14-layer convolutional neural network on raw single-channel EEG signals, deliberately selecting filter sizes to optimize detection.

Recurrent architectures explored interepoch transition modeling. Supratak et al. [65] created a hybrid model combining two convolutional neural networks with different filter sizes to capture characteristics in the time and frequency domain. A bidirectional LSTM was employed to learn transition rules between epochs. Zhang et al. [66] proposed a CNN with

a fast discriminative complex value, utilizing extended dimensionality to represent patterns more effectively, and applied the Winograd algorithm to reduce computational complexity.

Multimodal approaches investigated comprehensive signal integration. Mikkelsen et al. [67] developed a personalized sleep staging approach by training a deep learning model on combined EOG and EEG records. The general model was fine-tuned using individual subject data to capture personalized information. Another approach [68] introduced an end-to-end deep learning model for sleep stage classification using multivariate time series from EEG, EOG, and EMG signals, employing spatial filters and leveraging temporal context from neighboring time segments.

### 2.2.6 Combined Approaches

Hybrid methodologies integrated signal processing with representation learning to enhance interpretability and performance. Cecotti [69] proposed a convolutional neural network architecture that incorporates the Fourier transform to convert features in the time domain to the frequency domain. An innovative backpropagation approach transferred error between domains and a hybrid rejection strategy was introduced to increase the reliability and generalization of the model.

Recurrent neural networks used temporal information learning capabilities. Researchers explored various hybrid approaches, such as Hsu et al. [70], who utilized finite impulse response bandpass filters to separate characteristic brain waves and transformed signals using Fast Fourier Transformation. Another study by Yulita et al. [71] applied Deep Belief Networks in conjunction with Bi-LSTM on handcrafted features from EEG, EMG, and EOG signals, demonstrating superior performance compared to individual models.

Transfer learning and model adaptation strategies expanded the research frontiers. Vilamala et al. [72] applied multi-taper spectral estimation to single-channel EEG, converting spectrogram epochs to RGB color matrices and employing a pre-trained VGGNet model. Another approach

[72] developed a mapping algorithm to transfer feature representations learned from a well-annotated sleep staging database to different feature spaces, demonstrating the potential for the adaptation of the model from datasets.

The detailed exploration of these methodological approaches highlights the continuous innovation in sleep stage classification, showcasing the potential to integrate advanced signal processing and machine learning techniques.

## 2.3 Sleep Disorder Detection

This section explores the detection methodologies for major sleep disorders, with particular focus on obstructive sleep apnea (OSA), periodic limb movement syndrome (PLMS), and insomnia, using various sleep signal analysis techniques in multiple diagnostic frameworks and technological implementations.

### 2.3.1 Sleep apnea and periodic limb movement detection

Obstructive sleep apnea (OSA) represents one of the most prevalent conditions among the more than 70 sleep disorders cataloged by AASM [73]. According to AASM 2012 guidelines [56], OSA diagnosis requires episodes exceeding 10 seconds with airflow reduction of 90% (apnea) or 30% (hypopnea), accompanied by arousal or oxygen desaturation of at least 3%. The gold standards for diagnostics include ECG, oxygen saturation and nasal air flow measurements derived from PSG recordings [54]. Early detection of obstructive sleep apnea (OSA) has been demonstrated to mitigate risks of cognitive decline [74] and cardiovascular complications [75], making accurate and timely diagnosis crucial for patient outcomes.

The frequent co-occurrence of PLMS and OSA [76], linked to deeper resumption of breathing, necessitates simultaneous assessment in clinical settings. This complex interaction between respiratory and movement disorders presents unique challenges in diagnosis and treatment

planning. This section reviews research approaches classified by their methodological framework: signal processing or deep learning techniques, with particular attention to the evolution of detection methodologies and their clinical applications.

#### 2.3.1.1 Signal processing approaches

Recent investigations have extensively explored various signal processing methodologies for the detection of sleep disorders. In examining blood oxygen saturation (SpO2) signals, researchers [77] used least squares support vector machine classification in data from 100 patients with OSA, differentiating between those with and without cardiovascular comorbidity. This approach enabled a precise identification of the discriminative characteristics crucial for patient stratification. Similarly, [78] used fast correlation-based filtering in home-collected SpO2 recordings from 320 subjects, implementing multiple machine learning models to categorize levels of severity of sleep apnea, demonstrating superior performance compared to the conventional oxygen desaturation index 3%. The success of these home-based monitoring systems suggests promising directions for accessible diagnostic tools.

LS-SVM classification has shown versatility in different signal types. The research in [79] applied it to pulse photoplethysmographic (PPG) signal characteristics for OSA detection in 1-min segments from 26 PSG recordings, achieving significant accuracy in event detection. Meanwhile, [80] utilized it for OSA classification using single-lead ECG histogram features, demonstrating the potential of simplified monitoring approaches.

Heart rate analysis has emerged as a crucial metric in OSA detection, with multiple methodological approaches providing significant information. Studies have employed various sophisticated techniques: fuzzy sliding trend entropy for heart rate variability (HRV) analysis [81], which outperformed traditional time frequency domain analysis, time domain heart period analysis for nonstationarity investigation [82], and heart rate-adaptive match filtering for microvolt T wave alternans evaluation [83]. Notable research [84] established connections between PLMS and cardiovascular outcomes through HRV analysis, suggesting potential applications in home health monitoring using wearable ECG devices.

Combined detection of OSA and PLMS has been achieved through innovative approaches incorporating multiple sensing modalities. These include the PCA application on undermattress pressure sensors [85], which captured both breathing patterns and limb movements, real-time radar-based breathing pattern detection [86] achieving 90% accuracy in 100-hour experiments, and 3D time-of-flight camera monitoring [87] capable of detecting motor events beyond traditional EMG measurements. Advanced sensor systems have been developed, such as magnetometer-based breathing movement detection [88] offering energy-efficient alternatives to traditional methods, and multichannel analysis combining ECG and respiratory signals [89], [90] with improved feature selection methodologies.

Several comprehensive reviews have synthesized these developments, providing crucial insights into the effectiveness of methodology and clinical applications. Notable analyzes include impacts of the implementation of AASM [91], revealing significant variations in the classification of hypopnea and sleep stage, the relationships between RLS / PLMS and cardiovascular outcomes [17], [18] highlighting important health correlations, algorithmic performance evaluations [19] spanning 2003-2017, and developments in EMG signal analysis [20] encompassing approaches to temporal, frequency domain, and sparse representation.

#### 2.3.1.2 Deep learning approaches

Deep learning applications in the detection of sleep disorders, while fewer in number, have shown promising results with increasing sophistication in methodology and implementation. Temporal information capture has been achieved through the implementation of RNN with batch normalization [92], achieving 97.8% precision in OSA detection through innovative data augmentation techniques. CNN applications [93] have demonstrated superior performance compared to traditional SVM approaches in respiratory channel analysis, processing 24,480 30-second epochs with remarkable efficiency.

Comparative studies have evaluated various neural network architectures against traditional methods, providing comprehensive performance metrics and implementation insights. The research in [94] compared the k-nearest neighbor, self-organizing maps, and multilayer perceptron performance in the detection of ECG-based apnea, utilizing innovative Poincare

plot features. Furthermore, wearable sensor applications [95] have explored multiple machine learning approaches for PLM classification, addressing traditional PSG limitations through foot-mounted motion sensors.

### 2.3.2 Insomnia detection

Insomnia is the most common sleep disorder [96] and has significant implications for mental health, potentially leading to depression, anxiety [97] and cognitive impairment [98]. Its subjective nature and lack of standardized diagnostic criteria [99] have led to investigations into objective measures through EEG, heart rate variability [100], brain metabolite levels [101], and associated symptoms such as low back pain [99]. This multifaceted approach to diagnosis reflects the complex nature of the disorder and its varied manifestations.

The potential of the condition to induce serious health complications, including depression, diabetes, hypertension, and cardiovascular disease [102], underscores the importance of automated analysis systems for efficient diagnosis and monitoring. The necessity of comparing patient data with healthy subject recordings has driven the development of automated analysis tools to assist clinical diagnosis. Comprehensive reviews of patient-reported results [103] have highlighted the need to refine non-restorative sleep measurement methodologies, analyzing 26 studies that evaluated both the night and day aspects of non-restorative sleep.

Recent technological advances have enabled sophisticated detection approaches with increasing accuracy and reliability. Deep neural network applications [104] have shown success in EEG-based insomnia detection, using 57 temporal and spectral features of two-channel EEG recordings. Subsequent research [105] has explored computationally efficient alternatives using SVM with hypnogram features, indicating promising results with simplified models. These developments suggest a trend toward more accessible and efficient diagnostic tools while maintaining clinical accuracy.

## 2.4 Beyond Sleep Staging and Disorder Detection

Recent advances in sleep research have expanded beyond traditional sleep staging and disorder detection, encompassing diverse methodologies and applications. This section examines various research directions that, while not directly focused on sleep staging and disorder detection, have made significant contributions to our understanding of sleep-related phenomena and their practical applications in healthcare and daily life.

### 2.4.1 Signal Processing and Physiological Monitoring

The development of sophisticated signal processing techniques has revolutionized our ability to analyze and interpret physiological data related to sleep. These advances have particularly focused on the analysis of EEG signal and the integration of multiple physiological parameters for a comprehensive assessment of sleep.

In a groundbreaking study, [106] introduced an innovative approach that uses separable and depth-wise convolutions to capture the oscillatory and event-related features of EEG signals. Their methodology incorporated class weight adjustments in the loss function to address class imbalance issues, while employing one-way Analysis of Variance (ANOVA) for statistical validation of both within-subject and cross-subject models. The resulting features demonstrated remarkable capabilities for interpretable analysis through deep learning frameworks, establishing a new paradigm for EEG signal processing.

Building on these foundations, [107] developed sophisticated convolutional autoencoders (CAE) capable of learning individually adapted components with intersubject linkages. Their innovative approach used learned filter weights as initial values for subject-specific autoencoders, effectively capturing individual variations while maintaining cross-subject compatibility. The methodology incorporated a novel pretraining strategy for encoding relative similarity constraints, addressing fundamental limitations in previous encoder designs and enhancing the robustness of feature extraction.

Further advancing the field, [108] conducted a comprehensive comparison of neural network architectures, implementing a hierarchical approach that included a 2-layer "shallow" network, a 5-layer deep network, and a 31-layer residual network (ResNet), along with a hybrid CNN. Their statistical analysis compared decoding accuracies with the established filter bank common spatial patterns (FBCSP) algorithm, utilizing a cropped training approach with a 2-second sliding window to enhance accuracy. This study particularly highlighted the potential benefits of integrating the power features of the EEG band and the phase-related characteristics in deep convolutional neural networks.

In the realm of sleep disorder detection, [109] developed a portable REM sleep disorder detection system specifically designed for Alzheimer's diagnosis. Their system employed a linear SVM to classify REM and NREM states based on spectral edge frequency analysis, incorporating low to high spectral content evaluation and specialized REM detection algorithms. The implementation of a 16-point FFT enabled efficient hardware utilization while minimizing precision loss, demonstrating the practical applicability of advanced signal processing techniques in clinical settings.

### 2.4.2 Environmental Factors and Behavioral Analysis

The investigation of environmental influences on sleep quality has emerged as a crucial area of research, incorporating both technological innovations and traditional assessment methods to understand the complex relationships between external factors and sleep patterns.

A comprehensive study by [110] examined the differential impacts of pink-colored light versus simulated daylight on sleep quality, employing a multifaceted assessment approach that included Critical Flicker Frequency measurements, fatigue questionnaires, Karolinska sleeping scale evaluations, Pittsburgh Sleep Diary analysis and hormone concentration measurements. Their findings revealed that 2700K pink light could significantly enhance relaxation and improve overall sleep quality. Complementing these findings, [111] conducted detailed

investigations into the influence of lighting conditions on young people's sleep patterns, comparing various correlated color temperatures while simultaneously monitoring physiological parameters, including heart rate, blood pressure, and EEG signals.

In the domain of behavioral analysis, [112] pioneered innovative methods for fatigue assessment through social media analysis, extracting facial demographics from an extensive dataset of 1 million timeline posts from 10,480 Instagram users. Their research developed novel predictive models for individual fatigue rates based on facial landmark analysis, establishing important correlations between fatigue patterns and demographic factors. This work was further enhanced by [113], who developed an intelligent control system for electrical appliances based on sophisticated facial feature analysis for the detection of sleep state.

### 2.4.3 Clinical Applications and Healthcare Integration

The integration of sleep research into clinical practice has led to significant advances in patient care and monitoring systems, particularly in specialized healthcare settings and for specific patient populations.

A notable contribution in this area came from [114], who implemented a sophisticated Multimodal, Multiview Motion Analysis and Summarization (MASH) system for the monitoring of ICU patients. Their system utilized three RGB-D cameras and developed a modified Hidden Markov Model for sleep-pose pattern analysis, effectively addressing challenges in depth sensor noise reduction and motion quantization. This work was complemented by [115], who proposed an innovative spectrogram-based algorithm for real-time sleep position recognition using wearable sensors, achieving state-of-the-art accuracy through a long-term short-memory model.

In the field of sleep apnea diagnosis, [116] developed a comprehensive integrated system that combines multiple data sources, including clinical interviews, physical examinations, laboratory tests, PSG, and CPAP data. This holistic approach demonstrated the feasibility of integrating heterogeneous data sources for improved diagnosis and treatment planning. Further advancing clinical applications, [117] developed sophisticated Matlab-based tools for

quantitative and graphical presentation of hemodynamic variability, analyzing extensive ECG and impedance cardiographic signals from healthy subjects and patients with atrial fibrillation.

The relationship between sleep and memory consolidation was explored by [118], who designed an adaptive binary D optimal neural response model to optimize stimulation protocols based on sleep spindle and slow-wave oscillations. This research was particularly significant in establishing connections between sleep patterns and cognitive function. Additionally, [119] contributed to clinical diagnostics by developing a predictive model for OSA using selected features from the STOP-BANG questionnaire, analyzing a substantial clinical database of 856 PSG records to establish important correlations between OSA prevalence, gender, and age. Their rigorous statistical analysis revealed significant patterns in OSA prevalence across different demographic groups, providing valuable insights for clinical practice and treatment planning.

## 2.5 Public Datasets

This section presents a comprehensive overview of publicly available datasets suitable for sleep research and related physiological studies. The availability of public datasets has greatly advanced sleep research, allowing for reproducible studies and standardized benchmarking of analytical methods. These datasets vary in their scope, population demographics, and recorded parameters, making them valuable resources for different research objectives and methodological approaches.

### 2.5.1 General Sleep Research Datasets

The field of sleep research relies heavily on well-documented and thoroughly validated datasets. Three major databases have emerged as fundamental resources for general sleep research, each offering unique characteristics and research possibilities, as shown in Table 2.1.

**The Sleep-EDF Database** [120] represents one of the foundational datasets in sleep analysis research. It comprises horizontal EOG, Fpz-Cz and Pz-Oz EEG recordings sampled at 100Hz,

TABLE 2.1: General Sleep Research Datasets

| Dataset | Key Characteristics | Population |
|---|---|---|
| Sleep-EDF | • 100Hz EOG, Fpz-Cz and Pz-Oz EEG<br>• 1Hz EMG envelope data | Healthy Caucasians (21-35 years) |
| Sleep-EDF (expanded) | • 61 PSGs and hypnograms<br>• Additional measurements: submental chin EMG, oro-nasal respiration, rectal body temperature<br>• EDF+ compliant | • 79 healthy subjects (25-101 years)<br>• 22 subjects with mild sleep difficulties |
| MASS Database | • 256 Hz whole-night recordings<br>• EEG, EOG, EMG, ECG, respiratory signals<br>• Multicenter data from 3 laboratories | 200 participants:<br>• 97 males (23-63 years)<br>• 103 females (19-57 years) |

along with 1Hz sampled EMG envelope data. The subjects were healthy Caucasian males and females aged 21-35 years, with data collected in two phases: healthy volunteers in 1989 and subjects with mild sleep onset difficulties in 1994. The particular strength of this dataset lies in its clean baseline recordings from healthy subjects, making it especially valuable for fundamental sleep pattern research and algorithm development.

An expanded version, the Sleep-EDF database [expanded] (Sleep-EDF (x)) [121], was released in 2002 and is now recommended for contemporary research. This enhanced dataset encompasses 61 PSGs and hypnograms from 79 healthy subjects (aged 25-101) and 22 subjects with mild sleep onset difficulties. The expansion added crucial measurements including submental chin EMG, oro-nasal respiration, rectal body temperature, and event markers in *PSG.edf files, with all EDF header fields conforming to EDF+ specifications. The wider age range and additional physiological parameters make this dataset particularly suitable for aging-related sleep studies and multiparameter sleep analysis research.

**The Montreal Archive of Sleep Studies (MASS) Database** [122] serves as a comprehensive benchmark resource for automatic sleep analytics systems. It contains whole-night recordings

sampled at 256 Hz from 200 participants (97 males aged 23-63, 103 females aged 19-57), including EEG, EOG, EMG, ECG, and respiratory signals. The data originates from three different hospital sleep laboratories following eight distinct research protocols. This multicenter approach enhances the dataset's robustness and applicability across different clinical settings. The MASS database is particularly valuable for developing and validating automated sleep staging algorithms due to its diverse subject population and standardized recording protocols.

### 2.5.2 Specialized Sleep Disorder Datasets

Sleep disorders present unique challenges in both research and clinical practice. Several specialized datasets have been developed to address specific sleep-related conditions and their physiological manifestations. The St. Vincent's University Hospital / University College Dublin(UCD) Sleep Apnea Database [123] focuses on sleep-disordered breathing, containing 25 whole-night PSGs from adult subjects (21 men, 4 women; average age 50, weight 95kg, height 173cm). Each recording includes two EEG channels, two EOG channels, one EMG channel, and detailed hypopnea event annotations. The precise annotation of hypopnea events makes this dataset particularly valuable for developing automated detection algorithms and studying the temporal patterns of sleep-disordered breathing. The inclusion of anthropometric data also enables research on the relationship between physical characteristics and severity of sleep apnea. An overview of those datasets is illustrated in Table 2.2

**The Sleep Heart Health Study (SHHS) dataset** [124, 125, 126] is specialized in cardiovascular implications of sleep-disordered breathing. It comprises PSG recordings from 6441 subjects collected between 1995-1998 in unattended home settings, measuring EEG, EOG, ECG, EMG, nasal airflow, heart rate, fingertip pulse oximetry, thoracic and abdominal excursions, body position, and ambient light. The large-scale nature of this dataset, combined with its focus on home recordings, makes it invaluable for epidemiological studies and research into the relationship between sleep disorders and cardiovascular health. The unattended setting also provides insight into sleep patterns in natural environments, contrasting with laboratory studies.

TABLE 2.2: Specialized Sleep Disorder Datasets

| Dataset | Key Characteristics | Population |
|---|---|---|
| UCD Sleep Apnea | • Whole-night PSGs<br>• 2 EEG, 2 EOG, 1 EMG channels<br>• Detailed hypopnea annotations | 25 adults:<br>• 21 men, 4 women<br>• Avg: 50 years, 95kg, 173cm |
| Sleep Heart Health Study | • Unattended home recordings<br>• EEG, EOG, ECG, EMG, nasal airflow<br>• Heart rate, pulse oximetry<br>• Body position, ambient light | 6441 subjects (1995-1998) |
| CAP Database | • PSG recordings<br>• Multiple sleep disorders covered<br>• Includes control group | • 16 healthy subjects<br>• 92 pathological cases |

**The Cyclic Alternating Pattern (CAP) database** [127] contains 108 PSG recordings, including 16 healthy subjects and 92 pathological recordings from patients with various conditions including REM behavior disorder, nocturnal frontal lobe epilepsy, narcolepsy, insomnia, PLM, bruxism, and sleep-disordered breathing. The strength of this dataset lies in its comprehensive coverage of different sleep disorders, which allows comparative studies between multiple conditions. Inclusion of healthy control subjects allows research into the specificity and sensitivity of diagnostic criteria and the development of automated classification systems for sleep disorders.

### 2.5.3 Specialized Physiological Datasets

Beyond general sleep research, specialized datasets focus on specific physiological aspects and sleep-related phenomena, providing detailed insights into particular aspects of sleep physiology and related conditions. Those datasets are summarized in Table 2.3

**The Massachusetts Institute of Technology - Boston Dataset** [128] focuses on research on cardiac dynamics. It contains 48 thirty-minute excerpts of ambulatory two-channel ECG (360 Hz sampling rate) from 47 subjects, including both routine recordings and cases with

TABLE 2.3: MIT-Boston and DREAMS Datasets

| Dataset | Key Characteristics | Population |
|---|---|---|
| MIT-Boston Dataset | • 30-minute ECG excerpts<br>• 360 Hz sampling rate<br>• Normal and arrhythmic recordings<br>• Focus on cardiac dynamics | 47 subjects |
| DREAMS Subjects | • Whole-night PSG recordings<br>• R&K and AASM staging annotations | 20 healthy subjects |
| DREAMS Patients | • Multiple sleep disorders covered<br>• Comprehensive clinical annotations | 27 subjects with:<br>• Dysomnia<br>• RLS<br>• Insomnia<br>• Apnea/hypopnea |
| DREAMS Apnea | • Full overnight PSGs<br>• Annotated respiratory events<br>• Central, obstructive, mixed apnea | 12 sleep apnea patients |
| DREAMS PLMs | • Full overnight PSGs<br>• Detailed PLM annotations in EMG | 10 subjects with PLM |
| DREAMS Artifacts | • 15-minute excerpts<br>• Detailed artifact annotations | Data from 20 recordings |
| DREAMS Spindles | • 30-minute EEG segments<br>• Expert-annotated sleep spindles | 8 subjects |
| DREAMS K-complexes | • 30-minute EEG segments<br>• Expert-annotated K-complexes | 5 subjects |
| DREAMS REM | • 30-minute EEG segments<br>• Expert-annotated REM periods | 9 subjects |

significant arrhythmias [129]. The high sampling rate and careful selection of cases make this dataset particularly valuable for studying the relationship between sleep states and cardiac function. The inclusion of both normal and arrhythmic recordings enables comparative studies and the development of diagnostic algorithms.

**The DREAMS Database** [130], developed by the University of MONS - TCTS Laboratory, represents one of the most comprehensive and well-structured collections of sleep-related recordings. It comprises eight specialized databases, each focusing on specific aspects of sleep physiology and pathology:

- **The DREAMS Subjects Database:** Contains 20 healthy subjects' whole-night PSG recordings with both R&K and AASM staging annotations. This dual annotation system makes it particularly valuable for studying the differences between these two major sleep scoring standards and their implications for research and clinical practice.
- **The DREAMS Patients Database:** Encompasses 27 subjects with various pathologies including dysomnia, restless legs syndrome, insomnia and apnea/hypopnea syndrome. The diversity of conditions represented makes this database especially useful for differential diagnosis studies and the development of automated classification systems.
- **The DREAMS Apnea Database:** Features twelve full overnight PSG recordings from sleep apnea patients with expert-annotated respiratory events (central, obstructive, and mixed apnea and hypopnea). The detailed annotation of different types of respiratory events makes this dataset particularly valuable for studying the pathophysiology of sleep apnea and developing automated detection algorithms.
- **The DREAMS PLMs Database:** Includes ten full overnight PSG recordings with PLM annotations in EMG, providing essential data for studying periodic limb movements and their impact on sleep quality. The specific focus on PLM makes this dataset valuable for the development of automated detection systems and studying the temporal patterns of movement disorders during sleep.

- **The DREAMS Artifacts Database:** Contains twenty 15-minute excerpts with artifact annotations, crucial for developing robust signal processing algorithms and improving the quality of sleep analysis methods. The systematic documentation of artifacts helps in developing better filtering and preprocessing techniques.
- **The DREAMS Spindles Database:** Comprises eight 30-minute EEG segments with sleep spindle annotations, essential for studying these important markers of stage 2 sleep and their relationship to memory consolidation and cognitive function.
- **The DREAMS K-complexes Database:** Features five 30-minute EEG segments with K-complex annotations, providing valuable data for studying these distinctive elements of stage 2 sleep and their physiological significance.
- **The DREAMS REM database:** Contains 9 30-minute EEG segments with REM annotations, crucial for studying rapid eye movement patterns and their relationship to dream states and cognitive processes during sleep.

Each component of the DREAMS database serves specific research needs while maintaining high standards of data quality and annotation accuracy. The comprehensive nature of this collection, combined with its careful organization and documentation, makes it an invaluable resource for both focused studies on specific sleep phenomena and broader investigations of sleep physiology and pathology.

### 2.5.4 Dataset Limitations and Research Challenges

While the aforementioned datasets have significantly advanced sleep research, several limitations warrant consideration. Most datasets exhibit demographic biases, with overrepresentation of Caucasian populations and limited diversity in age, ethnicity, and socioeconomic backgrounds, potentially limiting the generalizability of research findings across diverse populations. Sample sizes for specific conditions remain modest, particularly for insomnia studies where subjective diagnostic criteria and recruitment challenges result in smaller cohorts compared to more easily identifiable disorders like sleep apnea. Additionally, many datasets lack longitudinal components, limiting insights into sleep pattern evolution over time, and environmental factors are often inadequately documented, despite their known influence

on sleep quality. These limitations highlight the need for more diverse, larger-scale, and comprehensive datasets to fully address the complexities of sleep medicine research.

## 2.6 Key Research Gaps and Challenges

Building on the comprehensive review of sleep signal analytics and the discussion of existing datasets and methodologies, several critical research gaps and challenges emerge that warrant further investigation. These gaps not only highlight the limitations of current approaches but also provide a foundation for the novel contributions presented in this thesis.

Table 2.4 summarizes the key research gaps and challenges addressed in this work.

TABLE 2.4: Summary of Key Research Gaps and Challenges

| **Research Gap** | **Description** |
|---|---|
| *Explainability in Deep Learning Models* | Existing models for sleep staging lack transparency in decision-making. Integration with signal processing theories could enhance interpretability while maintaining performance. |
| *Data Scarcity for Insomnia Detection* | Limited availability of labeled insomnia datasets hinders model development. Transfer learning from sleep staging datasets could improve detection accuracy. |
| *Incorporation of Sleep Microstructure* | Most sleep staging models ignore sleep microstructure (e.g., CAP), which provides critical information about sleep quality, particularly in NREM stages. |

These gaps directly inform the contributions of this thesis:

- **Chapter 3** addresses explainability by integrating signal processing with deep learning for sleep staging.
- **Chapter 4** tackles data scarcity using domain adaptation for insomnia detection from EEG signals.
- **Chapter 5** incorporates sleep microstructure into staging models using CAP features.

While this thesis makes significant progress toward addressing these challenges, several avenues for future work remain open:

- Developing more sophisticated hybrid approaches that combine traditional signal processing with modern deep learning techniques.
- Investigating novel sensing technologies for non-invasive sleep monitoring that can achieve clinical-grade accuracy while improving patient comfort.
- Enhancing existing algorithms to better handle individual variations in sleep patterns and environmental factors that affect signal quality.
- Exploring the complex relationship between sleep patterns and various health conditions, particularly in underrepresented populations.

The final chapter revisits this table to discuss how the thesis contributions address these gaps and outlines future research directions.

CHAPTER 3

# A Residual based Attention Model for EEG based Sleep Staging

---

In this chapter, inspired by the Hilbert transform which can extract relevant features from EEG waveforms, we propose a multi-scale deep architecture by decomposing an EEG signal into different frequency bands as input to CNNs. To model global temporal context, we utilize the multi-head self-attention module of the transformer model to not only improve performance, but also shorten the training time. In addition, we choose residual based architecture which makes training end-to-end. Experimental results on two widely used sleep staging datasets, Montreal Archive of Sleep Studies (MASS) and sleep-EDF datasets, demonstrate the effectiveness and significant efficiency (up to 12 times less training time) of our proposed method over the state-of-the-art.

## 3.1 Introduction

It is of great importance to monitor a person's sleep quality, as sleep takes up to one-third of a person's lifespan and sleep disorders often cause poor sleep quality as well as indicate other possible health issues. To assist diagnosis and treatment of sleep disorders, a standard assessment often incorporates EEG signals through polysomnography (PSG) which is the gold standard for most sleep studies and comprises other physiological measurements (breathing, oxygen saturations, heart rate and electromyography) [4]. EEG signals capture brain waves via electrodes placed on the scalp of a subject. They can be used to identify different sleep stages.

The concept of sleep stages and the rules of sleep staging were first introduced by Rechtschaffen and Kales (R&K) [50] in 1968. Typically, EEG signals are segmented into 30-second

epochs and classified into one of the 6 sleep stages: wakefulness (Wake), 4 non-rapid eye movement (Non-REM) stages (i.e., Stage 1 to Stage 4), and rapid eye movement (REM). In 2007, the American Academy of Sleep Medicine (AASM) revised and expanded the rules [51]. It defined 3 stages: Wake(W), Rapid Eye Movement (REM) sleep and Non-REM sleep. Specifically, Non-REM sleep can be further split into Non-REM1 (N1), N2 and N3 by merging Stage 3 and Stage 4 of R&K rules due to limited clinical or physiological difference could be found between Stage 3 and Stage 4.

Traditionally, sleep staging requires trained experts' visual scoring on 30-second epochs segmented from PSG recordings, which is a very tedious, time-consuming, and subjective process [53]. Further, spectral bands (alpha, beta, delta and theta waves) can be discriminated from EEG based on frequency bands and amplitudes [52], and the frequency of occurrence of each wave can be used to differentiate sleep stages. For instance, beta waves with frequencies ranging from 13 to 30 Hz and amplitudes from 1-5 $\mu$V can be observed at concentrated attention in REM stage while Wake stage can be characterized by alpha band of 8-13 Hz along with beta band of 13-30 Hz at low amplitude levels. In addition, the agreement between scorers is often less than 90% depending on the experience and fatigue the raters have [57]. Clinical scoring of PSG recordings is time-consuming and requires specialist training to accurate determine sleep EEG pathology. Therefore, automatic sleep staging would be ideal for more effective and efficient diagnosis of sleep disorders.

To provide solutions for automatic sleep staging systems, researchers used to choose signal processing techniques to extract frequency domain features and time domain features from raw EEG signals to discriminate stages of sleep because there has been a clear and theoretical explanation on spectrum analysis in the field of signal processing. In those approaches, raw signals from EEG are normally pre-processed by band-pass filters [58, 15, 23] to extract spectrum features before feeding them into machine learning models.

On the other hand, using fewer channels during sleep monitoring can alleviate obtrusiveness and reduce costs. Specifically, to explore the potential of single-channel based sleep staging, Stepnowsky *et al.* [57] compared automated scoring from a single lead EEG along with EOG which were derived from annotated PSG records and found that the automatic staging

algorithm was comparable to manual scoring on all stages. In [61], multi-scale principal component analysis was applied to remove signal noise from Pz-Oz signals EEG electrodes and then discrete wavelet transform was used to extract informative features from denoised signals. To reduce data dimensions, six different statistical features were decomposed and selected as an input of an ensemble model which adopted rotational support vector machine to classify different sleep stages. Although signal processing approaches can provide easily explainable results, it still requires domain expertise to extract meaningful features like K-complex waves in EEG epochs.

Recently, due to the promising progresses of deep learning in many pattern recognition tasks such as object recognition and detection, a number of deep learning based methods have been proposed for sleep staging. Aiming at providing personalized sleep staging, Mikkelsen *et al.* [67] trained a deep learning based model using electrooculography (EOG) and EEG signals. They then fine-tuned the general model with over sampled first-night data of a specific subject to learn personal information which improved scoring on second-night data. Aggarwal *et al.* in [131] proposed a method to classify sleep into 4 stages using nasal airflow signal for sleep apnea patients who were treated with Continuous Positive Air Pressure (CPAP) to see if the sleep efficiency was improved. They found using conditional random field as output layer could better capture global temporal context. Chambon *et al.* [68] introduced an end-to-end deep learning model to classify sleep stages with multivariate time series from multiple modalities (EEG, EOG, EMG) using spatial filters. Neighbouring epochs were fed into deep learning models simultaneously to capture temporal context which increased accuracy. Yuan *et al.* in [132] argued that the correlations among multivariate biosignals contain extra information which is useful in sleep staging, therefore proposed a fusion-based attention model to learn informative representations from multivariate PSG records.

Using deep learning on raw signals from single channel EEG has also become popular. In [63], signals of the current epoch along with its preceding two and succeeding two epochs were passed through a convolutional layer with 20 filters followed by a pooling layer. Then, the 20 filtered signals were reshaped into 2 dimension output to capture information across different time windows in the next layers. To efficiently handle the data imbalance issue

which is common in sleep staging, a class-balanced batch sampling was used in each epoch. In [64], two preceding and one succeeding epochs of the reference epoch were fed into a 14-layer convolutional neural network (CNN) with purposely selected filter sizes for each layer to detect sleep stages on raw single-channel EEG signal from the Sleep Heart Health Study (SHHS) dataset [124, 125, 126]. Another way to learn temporal context across epochs is using recurrent neural network such as LSTM. DeepSleepNet [65] was proposed to use two CNNs with different filter sizes to capture time-domain and frequency-domain characteristics respectively from a raw single-channel EEG while a bidirectional-LSTM (Bi-LSTM) was used against the concatenated output of those CNNs to learn temporal context among epochs. Note that DeepSleepNet outperforms the univariant version of the method proposed in [68] in terms of all metrics.

However, there are four major problems to be addressed in the deep learning based sleep staging literature:

- Most deep learning based models highly rely on hyperparameter tuning (especially in feature extraction part) which is challenging to extend or scale up.
- Oversampling is widely used to address data imbalance problem in automatic sleep staging solutions. However, it could increase training time and may not be the best option for datasets that have fewer than 10,000 samples which are used in existing studies [133, 134]. Shortening training process could help to experiment more complex models and speed up model development.
- Although temporal context is considered in some studies like [65, 135], the training has to be divided into two stages: pre-training and fine-tuning. This is because training a very deep neural network could inject gradient vanishing problem which makes the model converge to local optima or diverge. Furthermore, as the learning rate is set to a very low value during the fine-tuning process, it takes more time to reach optimum or may even not be able to reach optimum.
- The output of recurrent neural networks (RNN) such as long-short term memory (LSTM), which has been widely used to model temporal context across epochs in automatic sleep staging systems, heavily depends on preceding states. Thus,

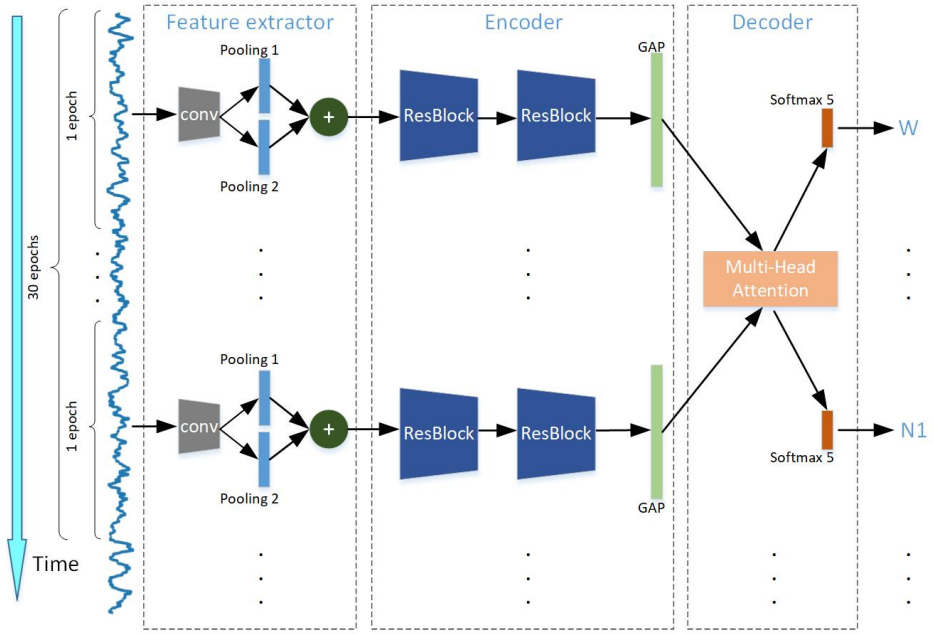


FIGURE 3.1: Overview of the proposed network architecture. There are three parts in the model: feature extractor, encoder, and decoder. The feature extractor is formed by a convolutional layer and a max pooling layer with two different scales. The encoder comprises of two residual blocks and a global average pooling (GAP) layer. The decoder consists of a multi-head attention module which models temporal context across 30-second epochs. A joint loss is used to optimize the classifier using the logits generated by both encoder and decoder.

the computation cannot be performed in parallel which also lengthens the training process.

In this chapter, we propose a novel residual based attention model using the raw single channel EEG signal to score sleep stages following the theory of Hilbert transform [136] which can represent both frequency and amplitude features. The main contributions of this work are as below:

- We adopt multi-scale max pooling layers to simulate Hilbert transform to extract frequency and amplitude features from EEG data.
- We incorporate 2 residual blocks which can be easily extended by stacking similar blocks to generate embeddings from extracted features for improved representation learning of different sleep stages.

- Instead of oversampling, we apply cost-sensitive learning which is proved more effective in small datasets to address data imbalance problem and speed up training.
- We propose a self-attention model to capture temporal context based on the multi-head attention mechanism in [137], which could be optimized in parallel to shorten training time significantly.
- To the best of our knowledge, our method is the first in the sleep staging literature to learn EEG epoch embeddings and temporal transition rule across epochs end-to-end because our model is based on residual network and the gradient vanishing issue can be completely avoided.

## 3.2 Proposed Method

An overview of our model's architecture is illustrated in Fig.3.1. We view sleep staging as a sequence-to-sequence task which transforms a sequence of extracted feature vectors from raw EEG signal into a sequence of sleep stages. Therefore, similar to the encoder-decoder structure which is frequently used in neural sequence transduction models, we designed three modules in our model: feature extraction, embedding generation (encoder) and temporal context capturing (decoder).

The feature extraction module extracts meaningful feature vectors via a convolutional layer followed by a max pooling layer with two different scales while the embedding generation module learns representation from those features and output embeddings of each epoch using two residual blocks followed by a global average pooling layer. The output sequence of generated embeddings are fed into a self-attention feed-forward neural network to capture temporal information across epochs.

During the optimization, the logits from embedding generation and from temporal context learning part are combined to infer the stage of an epoch as the final prediction through a joint loss.

### 3.2.1 Hilbert Transform and Feature Extraction

According to AASM, sleep stages are determined by frequency of occurrence of the different band waves. For instance, an alpha band of 8-13Hz and beta band of 13-30Hz are usually observed during the wakefulness stage while theta bands of 4-7Hz are dominant during N1 NREM sleep. As these wave bands are distinguished by their frequency and amplitude, to efficiently represent frequency and amplitude of different wave bands in 30-second EEG epochs, we found Hilbert transform in signal processing field could help describe those signals.

#### 3.2.1.1 Analytic signal and Hilbert transform

In signal processing theory, a real signal $s(t)$ (30-second epochs in sleep staging) can be represented as analytic signal [138]:

$$s_a(t) = s(t) + i \cdot H[s(t)], \tag{3.1}$$

where the real part denotes temporal context and the imaginary part is Hilbert transform of signal $s(t)$ which carries amplitude envelope information.

Hilbert transform of signal $s(t)$ can be defined as convolution of $s(t)$ with the signal $1/\pi t$ [139]:

$$H[s(t)] = \frac{1}{\pi t} * s(t), \tag{3.2}$$

which means Hilbert transform of a signal $s(t)$ equals feeding the signal into a bandwidth filter with an impulse response of $1/\pi t$.

#### 3.2.1.2 Module architecture

The design of the feature extraction module is as demonstrated in Fig.3.2.

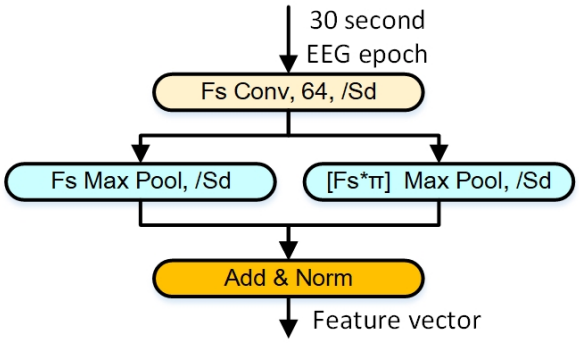


FIGURE 3.2: Design of feature extraction part. Fs and Sd are $\lceil N/100 \rceil$ and $\lceil N/1000 \rceil$ respectively where N denotes number of sampled points in an epoch. The two pooling layers of different sizes are meant to simulate Hilbert transform to extract frequency and amplitude features respectively.

Since the percentage (frequency of occurrence) of different waves in an epoch determines which sleep stage this epoch belongs to, we use a convolutional filter of size $\lceil N/100 \rceil$ in the first layer to extract meaningful features from raw signal, where $N$ denotes the number of data points in an epoch (30 seconds * sampling rate). Meanwhile, the stride is set to $1/10$ of the filter size which is $\lceil N/1000 \rceil$.

The output of the first layer is down-sampled by two max pooling layers with different pool sizes but same strides. The first pooling layer with size $\lceil N/1000 \rceil$ extracts temporal features like frequency (real part in Eqn.(3.1)) while the other pooling layer, which has a size of $\lceil \pi \times N/1000 \rceil$, is supposed to extract amplitude features (imaginary part in Eqn.(3.1)), .

To form up features for approximation of $\hat{s}_a(t)$ which is analytic signal of the input epoch, a point wise summation is applied to the output from those two pooling layers. Due to layer normalization [140] can speed up training in very deep neural network, it is applied on the output of all point wise summations in this model.

### 3.2.2 Residual-based Embedding Learning

To approximate mapping between input epoch $x$ and sleep stage $y$, we need multiple nonlinear layers as indicated in [141]. In the process of designing those layers, residual networks drew our attention.

#### 3.2.2.1 Residual network

Residual network (ResNet) [142] and its variants are popular approaches to solve image classification problems and whilst we did not identify empirical research from the literature in respect of sleep staging we have incorporated into our model. Typically, ResNet and its variants are established by multiple residual blocks which can be easily extended to very deep networks by simply stacking those residual blocks, so we decide to adopt residual blocks to learn the mapping under the hood.

ResNext is one of the best performed variants of ResNet till date. We choose aggregated residual transformation blocks from ResNext demonstrated in Fig.3(c) of [143] for residual learning following a split-transform-merge strategy which compresses network and saves computation resource.

A high level overview of embedding generation part is shown in Fig.3.3. Two residual blocks are used to generate embeddings from epochs. Below is an analysis on how those residual blocks approximate the mapping between input and output embeddings.

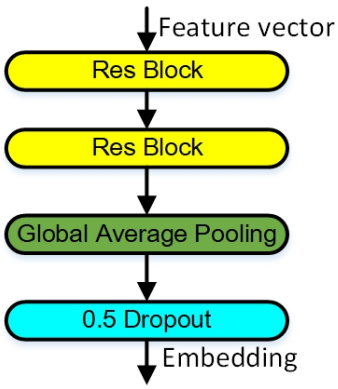


FIGURE 3.3: Embedding generation module is formed up with two ResNext blocks and a global average pooling layer.

Suppose there is a mapping $\mathcal{H}(X)$ between $X$ and $Y$:

$$Y = \mathcal{H}(X). \tag{3.3}$$

According to [142], instead of approximating $\mathcal{H}(X)$ directly, residual learning approximates the residual between $Y$ and $X$ as below:

$$\mathcal{F}(X) = Y - X = \mathcal{H}(X) - X. \tag{3.4}$$

Based on the above, it is possible to use residual units to approximate the function which transforms the extracted features to representation which is similar to analytic signal.

Let $x_i$ denote the input of the $i$th residual block and $\mathcal{F}_i$ denote the residual function which is supposed to be approximated by that block, then a typical residual block would follow the equation below:

$$x_{i+1} = \mathcal{H}_i(x_i) = \mathcal{F}_i(x_i) + x_i. \tag{3.5}$$

The function $\mathcal{F}_i(x_i)$ represents the residual mapping of the transform function $\mathcal{H}_i(x_i)$ which maps $x_i$ to $x_{i+1}$.

In our scenario, $x_1$ is the summation of frequency and amplitude features extracted from the prior module while the mapping between output of the first block $x_2$ and original input $x$ can be viewed as:

$$x \approx x_2 = \mathcal{H}_1(x_1), \tag{3.6}$$

where $x_2$ could be deemed as an approximation of analytic signal of $x$.

Now, the relationship between input epoch $x$ and output stage $y$ can be asymptotically approximated as below:

$$y = \mathcal{H}(x) \approx \mathcal{H}(x_2) = \mathcal{H}(\mathcal{H}_1(x_1)). \tag{3.7}$$

Apparently, we need another residual unit to learn the mapping $\mathcal{H}$.

#### 3.2.2.2 Module architecture

From above, we decide to use two ResNext blocks as shown in Fig.3.4.

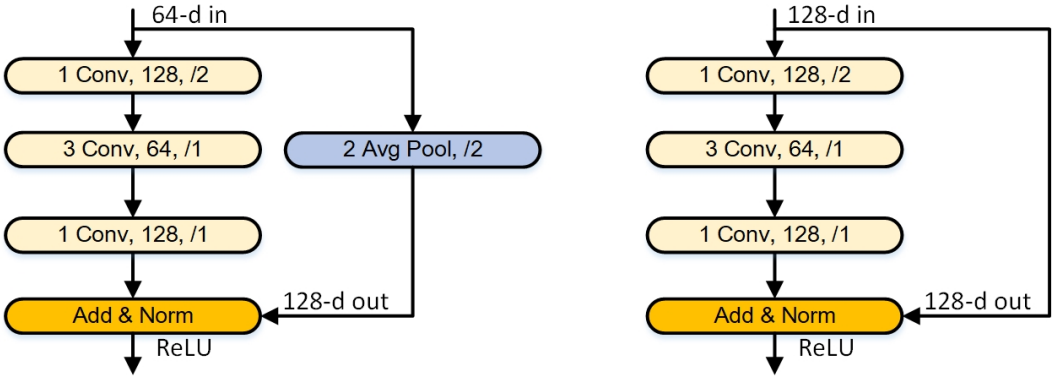


FIGURE 3.4: ResNext based residual blocks. The first block projects 64-dimension input into 128 dimensions and the second block maintains the same dimension with its input.

Since we use 64 filters in the first convolutional layer, it generates 64 feature maps for the first residual block. The input is split into lower-dimensional embeddings by $1 \times 1$ convolutions with a stride of 2 and then transformed by a grouped convolutional [144] layer consists of 64 $3 \times 1$ filters which concatenates the reformulations as input of next layer. Since we need to look at percentages of different waves in an epoch, the last layer of the residual block is formed up with 128 $1 \times 1$ filters to extract more than 100 feature maps and project the features learned in lower-dimensional feature space back to higher-dimensional embeddings.

Given the output dimension of the first residual block is 128 while the input dimension is 64, a projection shortcut connection (done by average pooling layer) is adopted to match dimensions. An element wise addition is applied upon residual learned by residual block and down-sampled original input to get the output $x_2$. A rectified linear unit (ReLU) activation is used on the summation.

The second residual block has the same input/output channels (128/128) to fully represent those extracted feature maps, so the identity shortcut connection is adopted in this block instead of a projection shortcut connection. We also experimented 64/64 input/output dimension block as the first block, however, with training time increased due to no down sampling, no performance gain was observed. On the other hand, we observed a performance deterioration when using 128/256 input/output block as the second residual block. This appeared be due to another average pooling layer causing information loss across two blocks. Similar to the first residual block, a ReLU activation is applied on the output.

A global average pooling layer [145] is used to average the feature maps obtained from residual blocks to generate embeddings as output to the next module.

### 3.2.3 Temporal Modeling with Self-attention

The stage of an epoch could be related to some preceding or succeeding epochs, which means the temporal information could be important for sleep scorers to determine a specific epoch. For instance, according to AASM, low amplitude and mixed frequency EEG pattern denotes N1 while at least one non-arousal K-complexes or no less than one trains of sleep spindles can identify N2, but if a subject is observed to enter sleep stage N2, epochs without K-complexes and sleep spindles should be annotated as N2 if they contain low amplitude and mixed frequency EEG activity.

To capture temporal context across different epochs, one of the classic approaches is using RNN [65, 71, 135]. Specifically, bidirectional RNN like Bi-LSTM is widely used to learn the transition rules from prior stages and also posterior stages. However, the computation cannot be processed in parallel due to a sequence of hidden states has to be generated in each step of time series which makes training process slow.

#### 3.2.3.1 Transformer

To efficiently capture temporal context across a sequence of PSG epochs, we explored in natural language processing (NLP) domain which has a number of research on replacing

LSTM in sequence-to-sequence tasks like translation. Among those solutions, we found an attention model [137] could be adopted to capture temporal context in sleep staging.

The model is called Transformer which outperforms LSTM in a lot of language translation tasks. As our task, sleep staging, can also be deemed as translate 30-second epochs into a sequence of stages, we build a self-attention model based on Transformer for time variant information learning and it shortens training process significantly because computation can be performed in parallel.

Given Transformer discards recurrence, to incorporate the order of the sequence, we encode the position of time steps in the sequence of input embeddings with sinusoidal functions, because parameterized position encoding does not improve performance according to [137]. To add position encoding to input embeddings from residual network, we use same dimension with the output from embedding generation module, which is 128. During experimentation, we find the model yields best performance when we feed 30 epochs in each sub-sequence, thus, the equation for the position encoding can be formularized as below:

$$PE_{(p,2d)} = \sin(p/10000^{2d/dim}), \tag{3.8}$$

$$PE_{(p,2d+1)} = \cos(p/10000^{2d/dim}), \tag{3.9}$$

where $p$ denotes the position of an epoch in the input sequence, $d$ is 30 which represents the length of sub-sequence and $dim$ is 128 as mentioned previously. Since the position embedding of an epoch can be calculated directly from above equation without training, it speeds up training and helps algorithm to learn the transition rule from both directions.

#### 3.2.3.2 Multi-head attention

In attention mechanism, the query $q$, which typically maps a set of key-value pairs $\{k, v\}$ to ground truth $y$, could be approximated by an attention function. The weighted sum of values, where the weights are calculated from a compatibility function, can be assumed as a

prediction of output. To reduce computation, we use self-attention which means the query and key-value pairs all correspond to the input sequence. Then, the attention function can be mathematically defined as below according to [137]:

$$Attention(X, X, X) = softmax(\frac{XX^T}{\sqrt{dim}})X, \tag{3.10}$$

where $X$ denotes matrix formed up by input embedding vectors and $dim$ is the dimension of those vectors.

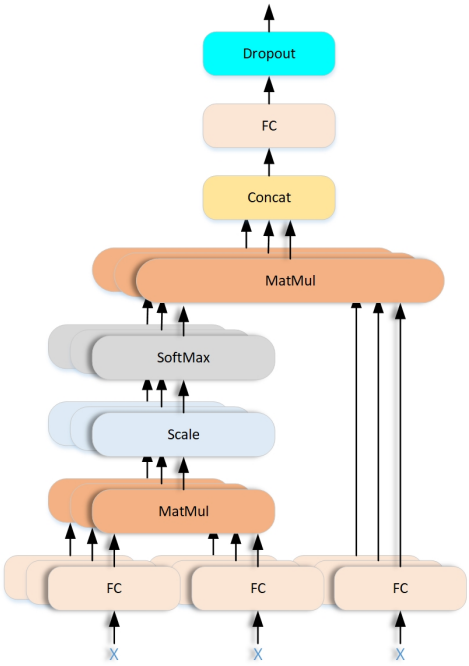


FIGURE 3.5: Multi-head self-attention sub-module. The sequence of embeddings is linearly projected 8 times in parallel and the output values are concatenated.

To have the model jointly attend to information from different representation subspaces at different positions thereby improving performance, we incorporate multi-head attention mechanism which linearly projects queries, keys and values (all of them are embedding sequences generated by previous module) $h$ times.

During hyper-parameters tuning, the best performed settings are found exactly same as those proposed in [146]. We use 8 for $h$ which is number of heads. In each head, the dimension is 64 and attention function is performed on each head in parallel as shown in Fig.3.5. The

denominator $\sqrt{dim}$ in Eqn.(3.10) is reflected as a scale layer and the output of each head is concatenated. Then, a fully connected layer is used to project the concatenated result from higher dimension ($8 \times 64 = 512$) to the same dimension of the original input embedding which is 128.

To regularize this sub-module, a dropout layer with keep rate of 0.8 is added before outputting the result to next module.

#### 3.2.3.3 Module architecture

Similarly to the encoder-decoder structure which is frequently used in neural sequence transduction models, we consider the previous embedding generation module as encoder and the encoded embedding sequence is decoded into stages by this temporal context learning module.

The overall design of this module is as shown in Fig.3.6. From it, we can see that the encoded embedding sequence is decoded by the multi-head attention sub-module along with two fully connected layers.

A residual connection is applied to add the input embedding and output from multi-head attention. A fully connected layer with 2048 neurons is used to project the attention output to high-dimensional representation and then the result is activated by ReLU. After that, another fully connect layer projects the activations back to 128 dimension. The normalized output from multi-head attention is also added to output of two fully connected layers by a skip connection.

To improve performance and avoid gradients vanishing issue during back propagation, the final output of the module is added to the original encoded embeddings via a shortcut connection.

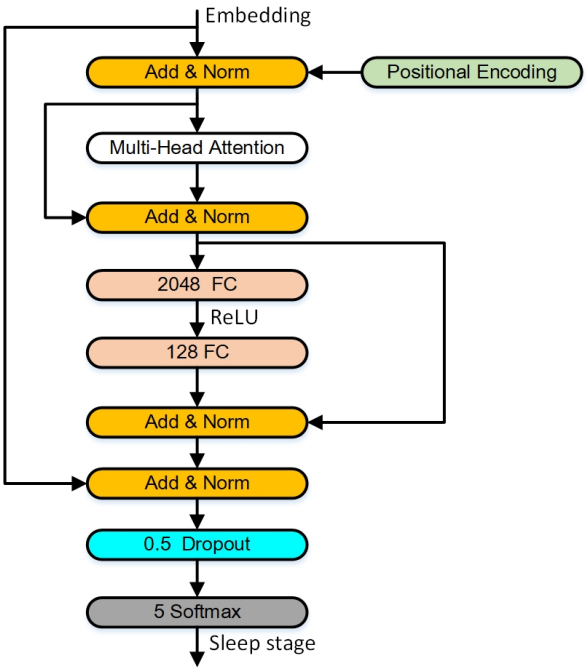


FIGURE 3.6: Temporal context learning module. Embeddings from previous module is fed into a multi-head attention network followed by two fully connected layers.

### 3.2.4 Training

To successfully train this network at speed, data imbalance problem is the first thing to be addressed. It is common in sleep staging and is usually managed by oversampling [65] or class balanced sampling [64, 147].

Oversampling, which clones samples of minority classes, lengthens training time as well as increases computation while class balanced sampling, which samples same amount of observations from different classes, could leave out some samples from predominate classes thereby sacrificing performance in representation learning.

#### 3.2.4.1 Cost-sensitive learning

To ensure all observations are used without lengthening the training, we decide to adopt cost-sensitive learning in our model.

Cost-sensitive learning which assigns weights as penalty for those less accurately classified samples could be more helpful in avoiding class imbalance issue in datasets that have less than 10,000 samples like EEG according to [134, 148, 133]. In addition, less training time and computation comparing with oversampling could be benifical to design more complex models which renders better performance.

Since our model removes oversampling and all sub-modules are connected by skip connections, which ensures gradients back propagate properly, we achieve end-to-end training on this very deep neural network using a joint loss.

#### 3.2.4.2 Multi-task learning

Inspired by multi-task learning (MTL) [149], we treat embedding encoding and temporal information capturing as two related tasks. Jointly training these two tasks can help they look at each other's latent features which could improve performance and generalization.

Firstly, we use a softmax layer to generate logits from output of encoding part and calculate cross-entropy loss [150] of $N$ classes (N equals 5 in this work which are Wake, REM, N1, N2 and N3) classification as below:

$$\mathcal{L}_c = -\sum_{c=1}^{N} y_c \log(y'_c), \tag{3.11}$$

where $y_c$ is true probability an observation belongs to class $c$ and $y'_c$ is the predicted probability of the observation. To apply cost-sensitive learning, we assign weights according to which class the observation belongs to. The class-weighted cross-entropy loss function is expressed as following equation:

$$\mathcal{L}_{wc} = -w_c \sum_{c=1}^{N} y_c \log(y'_c), \tag{3.12}$$

where $w_c$ denotes the weight of each class's contribution to final loss. The more samples from a class, the less weight the class will be assigned.

It can be derived from inverse of proportion of observations from different classes:

$$p_c = \frac{\sum_{c=1}^{N} S_c}{S_c}, \tag{3.13}$$

where $S_c$ denotes the number of samples from class $c$. Thus, $w_c$ can be calculated as:

$$w_c = N \frac{p_c}{\sum_{c=1}^{N} p_c}. \tag{3.14}$$

The average cross-entropy loss of encoding part's output is viewed as:

$$\mathcal{L}_e = \frac{1}{M} \sum_{i=1}^{M} \mathcal{L}_{wc,i}, \tag{3.15}$$

where $M$ is the size of mini-batch which is 300 in our work and $i$ is the location of the sample.

The output of encoding part is fed into transition rule learning part as mini-batches with 10 sub-sequences of length 30 in each batch. To optimize this module, the loss function is viewed as summation of cross-entropy loss across all samples in the sub-sequence:

$$\mathcal{L}_s = \sum_{i=1}^{O} \sum_{j=1}^{P} \mathcal{L}'_{e,(i,j)}, \tag{3.16}$$

where $\mathcal{L}'_{e,(i,j)}$ denotes the cross-entropy loss of $j$th sample in $i$th sub-sequence of the final output. $O$ is the size of mini-batch in temporal learning network and $P$ is length of a sub-sequence. They are 10 and 30 respectively in our model.

To make loss from encoding part and temporal learning part at same scale, we use a multi-task loss function in this form:

$$\mathcal{L}_{mtl} = \frac{1}{M} \mathcal{L}_s + \mathcal{L}_e, \tag{3.17}$$

where $M$ is 300 which is the same with that in Eqn. (3.15).

## 3.3 Experimental Results and Discussions

In this section, we report the results of experiments conducted on two popular EEG based sleep staging datasets for evaluating the effectiveness and efficiency of our proposed method. We first describe the two datasets: Montreal Archive of Sleep Studies (MASS) [122] and sleep-EDF [120], evaluation metrics, and experimental settings, then present experimental results and our conclusions.

### 3.3.1 Datasets

**MASS** contains whole-night recordings of EEG, EOG, EMG, ECG and respiratory signals from 97 males aged from 23 to 63 and 103 females aged from 19 to 57 years. Since these recordings were collected from 3 different sleep laboratories using 8 different research protocols, this dataset can be used as a centralized resource for solving difficult topics like improving inter-rater agreement on sleep staging or the impact of age on sleep. The signals in this dataset were recorded at a sampling rate of either 256Hz or 512Hz. To ensure the samples have the same dimension, the signals recorded at a sampling rate of 512Hz were downsampled uniformly. The recordings are in the European Data Format (EDF)++ and organized into five subsets (named SS1-SS5) according to the research protocols. SS1 and SS3 are annotated according to AASM while the others follow R&K rule. Since SS1 has a mixed montage size, we choose SS3 from cohort 1 which contains 28 male and 34 female healthy subjects aged from 20 to 69 years old. The epochs annotated as UNKNOWN are removed and the rest of the epochs are labeled into five sleep stages: W, N1, N2, N3 and REM.

The F4-LER (linked-ear reference with a 10 $k\Omega$ resistance) channel is chosen for our experiments by default. However, the F4-CLE (computed linked-ear) channel is the alternative when F4-LER does not exist in a recording. The reason that not use F4-EOG channel as in [65] and [151] is because we assume F4-EOG which is reformatted signal by subtracting the EOG channel to the EEG channel (e.g., F4-LER - (EOG-LER) = F4 - EOG - LER + LER = F4-EOG) may inject data leaking from EOG.

TABLE 3.1: The epoch statistics of MASS and sleep-EDF.

| Dataset | W | N1 | N2 | N3 | REM | Total |
|---|---|---|---|---|---|---|
| MASS | 6,231 | 4,814 | 29,777 | 7,653 | 10,581 | 59,056 |
| sleep-EDF | 8,285 | 2,804 | 17,799 | 5,703 | 7,717 | 42,308 |

**Sleep-EDF** is another popular dataset for sleep analysis, which consists of horizontal EOG, Fpz-Cz and Pz-Oz EEG at a sampling rate of 100Hz along with an EMG envelope at a sample rate of 1Hz. The signals were collected from 197 healthy Caucasian males and females aged from 21 to 35. To conduct fair comparisons with the literature, we used Fpz-Cz and Pz-Cz channels of recordings collected from 20 healthy subjects of the study on age effect in the Sleep-EDF dataset. Each epoch was scored into 8 classes according to the R&K rule: W, S1, S2, S3, S4, REM, MOVEMENT, and UNKNOWN. To be consistent with MASS, stages S3 and S4 were merged as N3 and and the epochs annotated as MOVEMENT or UNKNOWN were discarded. In addition, continuous wake epochs longer than 30 minutes outside sleep period were also ignored in our experiments.

The statistics of these two dataset are shown in Table 3.1.

### 3.3.2 Evaluation Metrics

Three evaluation metrics are used in our experiments: macro-averaging F1-score (MF1), overall accuracy (ACC), and Cohen's Kappa coefficient ($\kappa$). MF1 and ACC can be obtained as follows:

$$ACC = \frac{\sum_{i=1}^{K} TP_i}{N}, \tag{3.18}$$

$$MF1 = \frac{\sum_{i=1}^{K} F1_i}{K}, \tag{3.19}$$

where $TP_i$ and $F1_i$ are the true positives and F1-score of the $i$th stage respectively, $K$ is the total number of sleep stages (i.e., 5 in our study), and $N$ is the total number of test epochs.

Cohen's Kappa Coefficient [152] is to characterize the inter-rater agreement (IRA) level and can be calculated as following:

$$\kappa = \frac{N_c - N_e}{N_t - N_e}, \tag{3.20}$$

where $N_c$ denotes the number of correctly scored stages, $N_t$ denotes the total number of stages while $N_e$ denotes the expected number of agreements for each stage. The correspondence between Cohen's Kappa Coefficients and their agreement levels are shown in Table 3.2.

TABLE 3.2: The correspondence between different agreement levels and Cohen's Kappa Coefficients

| Cohen's Kappa Coefficient | Agreement Level |
|---|---|
| $\kappa$ <0.00 | Poor |
| $0.00 \leq \kappa \leq 0.20$ | Slight |
| $0.21 \leq \kappa \leq 0.40$ | Fair |
| $0.41 \leq \kappa \leq 0.60$ | Moderate |
| $0.61 \leq \kappa \leq 0.80$ | Substantial |
| $\kappa$ >0.80 | Excellent |

### 3.3.3 Experimental Settings

We split the MASS and sleep-EDF datasets into 31 and 20 folds respectively, In each fold, two recordings are considered as test data and the rest are used as training data. In particular, the two recordings are from the same subject in sleep-EDF. This process is repeated until all the folds have been iterated.

We used the Adam optimizer at a learning rate of $10^{-4}$ against the loss function defined in Eqn.(3.17). Since in our proposed model, there is no pre-training phase, which is essentially an initializing process, it is important to have a good initialization during training. Therefore, we adopt the initialization method proposed in [153], which produces the best performance.

To prevent the network from overfitting to noises or artifacts, L2 weight decay is applied to the first convolution layer of the encoding part with weight decay parameter set to $10^{-3}$. We also introduce two dropout layers in this model to make it more generalized. The first one is after global average pooling layer in embedding generation part while the other is prior to softmax layer which generates final output as shown in Fig.3.3 and Fig.3.6 respectively.

### 3.3.4 Comparison of Efficiency

We compared our proposed method with the state-of-the-art deep learning based sleep staging models including DeepSleepNet [65] and a ResNet based pure CNN model proposed by Ahmed *et al.* [154] in the same environment. For these models, we saved the last checkpoint for evaluation.

Our proposed method is implemented with TensorFlow [155] and the training is conducted on a computer with a GPU NVIDIA GeForce GTX 1080Ti. Our proposed model is trained end-to-end and generally converged in 100 epochs. Each fold takes less than 15 minutes for training, while DeepSleepNet and method in [154] take around 2.5 hours and 10 hours respectively. The inference or testing time of each fold is around 0.5 second. Both training and inference are more than 10 times faster than the other two approaches.

### 3.3.5 Computational Cost Analysis

To provide a comprehensive evaluation of our proposed method's efficiency, we present a detailed computational cost analysis comparing our approach with state-of-the-art methods.

**Model Complexity:** Our proposed model contains approximately 2.1M parameters compared to DeepSleepNet's 3.8M parameters and the ResNet-based method's 5.2M parameters. The reduction in parameters is achieved through the efficient multi-scale feature extraction and the use of global average pooling instead of fully connected layers.

**Memory Requirements:** During training, our model requires approximately 1.2GB of GPU memory per batch, which is 40% less than DeepSleepNet (2.0GB) and 60% less than the ResNet-based approach (3.1GB). This memory efficiency is primarily due to the absence of recurrent connections and the parallel processing capability of the self-attention mechanism.

**Training Time Breakdown:** The significant speedup (10-12x faster) can be attributed to three factors: (1) elimination of the two-stage training process (pre-training + fine-tuning) required by DeepSleepNet, (2) parallel computation in the self-attention mechanism vs.

sequential processing in LSTM, and (3) cost-sensitive learning eliminating the need for data oversampling.

**Inference Efficiency:** Our model achieves real-time processing with an average inference time of 0.02 seconds per 30-second epoch on a single GPU, making it suitable for clinical deployment scenarios.

### 3.3.6 Performance Comparison on Sleep-EDF

We trained our model on Fpz-Cz and Pz-Oz channels from Sleep-EDF dataset as well as F4 channel from MASS dataset. Per-class metrics like precision (PR), recall (RE) and F1-score (F1) are compared between proposed model and baselines side by side. Those metrics are calculated by treating the candidate class as positive and other classes as negative.

Tables 3.3 and 3.4 show confusion matrices obtained from the Fpz-Cz and Pz-Oz channels, respectively, of the sleep-EDF dataset for our model and baselines. Each row represents ground truth while each column indicates our predictions. The numbers of epochs which are correctly classified by our model are in bold. Also, per-class metrics from proposed model and baselines are appended in the last 6 columns with winner of each metric highlighted. We also compared overall accuracy (ACC), macro-averaging F1-score and Cohens Kappa in Tables 3.5 and 3.6 across the three approaches.

From those confusion matrices, our proposed method outperforms DeepSleepNet and method from [154] for both Fpz-Cz and Pz-Oz channels on Sleep-EDF in terms of almost all the three evaluation metrics (F1, precision and recall), which shows the effectiveness of our proposed method.

To demonstrate the statistical significance of improvement, we apply K-fold cross-validated paired t-test on each fold's accuracy for three methods. The p-values on Fpz-Cz channel are 0.0504 and 0.0548 comparing our proposed model to DeepSleepNet and method from [154] respectively while corresponding p-values on Pz-Oz channel are 0.1086 and 0.2134.

Although on Pz-Oz channel the performance of our proposed method is slightly lower than DeepSleepNet in REM stages, our proposed method clearly outperforms the baseline on other stages. This could be explained that Pz-Oz is placed at the centre of head which is far away from eyes and cannot capture the electrical activities which are influenced by movement of eyes. This causes misclassifications between stage W and N1. Model performs better on Fpz-Cz channel than on Pz-Oz channel because Fpz-Cz is placed on front head which is close to eyes.

We visualized annotated stages along with predicted stages of Fpz-Cz and Pz-Oz channels in Fig.3.7 and Fig.3.8 using the first subject"s recording from sleep-EDF. We can see that it is harder for the candidate to discriminate REM and W using signals from Pz-Oz than Fpz-Cz which could be also due to the distance between electrode and eyes.

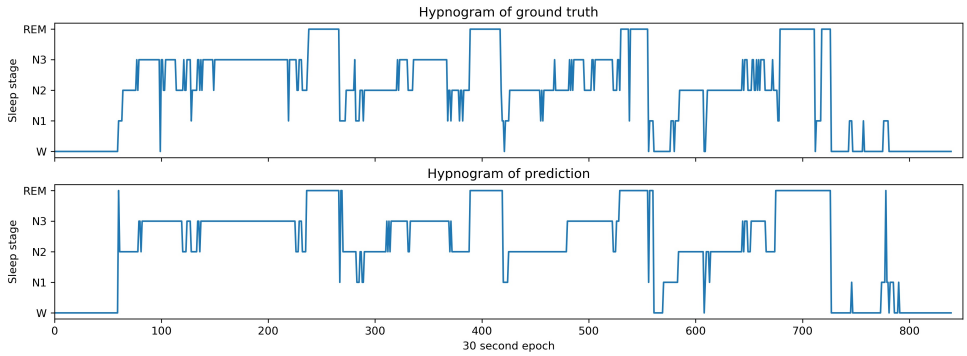


FIGURE 3.7: Hypnograms of subject 1 from sleep-EDF dataset Fpz-Cz channel. The top is ground truth and the bottom is prediction from our model. The accuracy and F1-score of our model are 0.90 and 0.83 respectively for this subject.

TABLE 3.3: Confusion matrix from 20 fold cross-validation of sleep-EDF Fpz-Cz channel and per-class metrics

| | Predicted | | | | | Candidate | | | SOTA Ref. [65] | | | Ref. [154] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 | RE | PR | F1 | RE | PR | F1 |
| W | **6604** | 229 | 38 | 5 | 36 | **95.5** | **85.4** | **90.2** | 88.8 | 83.4 | 86.0 | 86.9 | 85.0 | 85.9 |
| N1 | 720 | **1373** | 282 | 0 | 510 | **47.6** | **49.0** | **48.3** | 42.8 | 47.2 | 44.9 | 39.2 | 38.2 | 38.7 |
| N2 | 199 | 818 | **15860** | 571 | 900 | 86.4 | **89.1** | **87.8** | **88.2** | 82.3 | 85.1 | 85.6 | 83.2 | 84.4 |
| N3 | 33 | 17 | 1075 | **5125** | 16 | **81.8** | 89.9 | **85.6** | 76.8 | 92.7 | 84.0 | 74.3 | **92.8** | 82.5 |
| REM | 181 | 367 | 544 | 2 | **6255** | **85.1** | 81.1 | **83.0** | 81.5 | **83.7** | 82.6 | 80.2 | 73.2 | 76.5 |

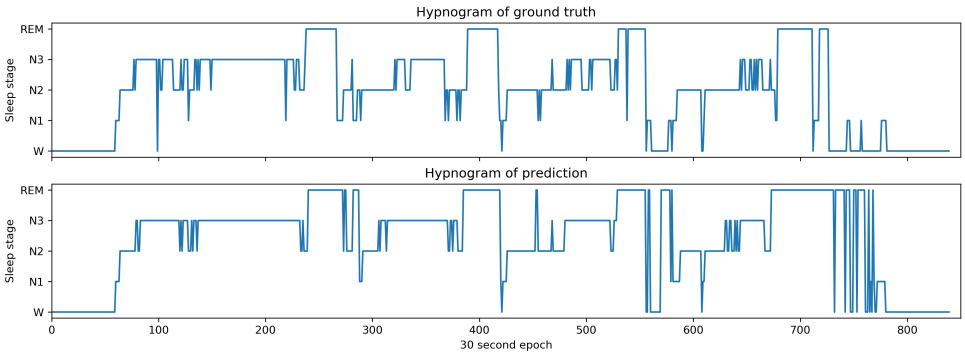


FIGURE 3.8: Hypnograms of subject 1 from sleep-EDF dataset Pz-Oz channel. The top is ground truth and the bottom is prediction from our model. The accuracy and F1-score of our model are 0.81 and 0.73 respectively for this subject.

TABLE 3.4: Confusion matrix from 20 fold cross-validation of sleep-EDF Pz-Oz channel and per-class metrics

| | Predicted | | | | | Candidate | | | SOTA Ref. [65] | | | Ref. [154] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 | RE | PR | F1 | RE | PR | F1 |
| W | **6672** | 544 | 139 | 13 | 205 | **88.1** | 86.2 | **87.2** | 86.3 | **87.0** | 86.7 | 87.7 | 86.6 | 87.1 |
| N1 | 659 | **1051** | 708 | 1 | 494 | 36.1 | **37.5** | **36.8** | **47.4** | 29.0 | 36.0 | 37.2 | 25.2 | 30.1 |
| N2 | 101 | 592 | **14856** | 885 | 660 | **86.9** | 83.5 | **85.2** | 82.5 | 82.4 | 82.5 | 82.2 | **84.4** | 83.3 |
| N3 | 25 | 25 | 1210 | **4784** | 16 | **78.9** | 83.9 | **81.3** | 74.3 | 71.4 | 72.8 | 71.8 | **88.0** | 79.1 |
| REM | 280 | 592 | 886 | 20 | **6342** | 78.1 | 82.2 | 80.1 | 75.2 | **87.5** | **80.9** | **80.1** | 72.2 | 75.9 |

TABLE 3.5: Comparison between our proposed method and state-of-the-art, using the Fpz-Cz channel of Sleep-EDF. The best performance in each column is highlighted in bold.

| | Overall Metrics | | | Per-class F1 | | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | MF1 | $\kappa$ | W | N1 | N2 | N3 | REM |
| SOTA Ref. [65] | 81.8 | 76.5 | 0.76 | 86.0 | 45.0 | 85.1 | 84.0 | 82.6 |
| Ref. [154] | 80.1 | 73.6 | 0.73 | 85.9 | 38.7 | 84.4 | 82.5 | 76.5 |
| Candidate | **84.3** | **79.0** | **0.78** | **90.2** | **48.3** | **87.8** | **85.6** | **83.0** |

### 3.3.7 Performance Comparison on MASS

Similar results are also observed on MASS dataset. As shown in Table 3.7, our proposed method is able to achieve comparable performance with DeepSleepNet (p-value from K-fold cross-validated paired t-test is 0.3384), while being significantly more efficient.

TABLE 3.6: Comparison between proposed model and the state-of-art across overall accuracy (ACC), macro-F1 score (MF), Cohen's Kappa ($\kappa$) on Pz-Oz channel from Sleep-EDF dataset

| | Overall Metrics | | | Per-class F1 | | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | MF | $\kappa$ | W | N1 | N2 | N3 | REM |
| SOTA Ref. [65] | 79.2 | 71.8 | 0.72 | 86.7 | 36.0 | 82.5 | 72.8 | **80.9** |
| Ref. [154] | 79.2 | 71.1 | 0.71 | 87.1 | 30.1 | 83.3 | 79.1 | 75.9 |
| Candidate | **80.7** | **74.1** | **0.74** | **87.2** | **36.8** | **85.2** | **81.3** | 80.1 |

TABLE 3.7: Comparison between proposed model and the state-of-art across overall accuracy (ACC), macro-F1 score (MF), Cohen's Kappa ($\kappa$) on F4 channel from MASS dataset

| | Overall Metrics | | | Per-class F1 | | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | MF | $\kappa$ | W | N1 | N2 | N3 | REM |
| SOTA Ref. [65] | **86.7** | **81.2** | 0.80 | **87.5** | **55.4** | 91.3 | 84.8 | **87.2** |
| Candidate | 86.5 | 81.0 | 0.80 | 87.2 | 52.9 | **91.5** | **87.0** | 86.6 |

It is noticed that model achieves higher performance on MASS than on Sleep-EDF in terms of all the three evaluation metrics. This could be explained that the MASS dataset has more samples and higher resolution to better train the deep networks.

We also show the confusion matrix in Table 3.8. From these confusion matrices, we can see the candidate performs worst in N1 stage (F1 ranges from 36.8 to 52.9). N1 epochs are likely to be misclassified as N2.

In addition, looking at per-class metrics on F4 channel from MASS, we can see proposed model outperforms baseline in classifying stage N2 and N3 while the baseline is good at distinguishing W, N1 and REM. It implies our model captures temporal transition rule better and the baseline is doing well in semantic representation learning.

TABLE 3.8: Confusion matrix from 31 fold cross-validation of MASS F4 channel and per-class metrics

| | Predicted | | | | | Candidate | | | SOTA Ref. [65] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 | RE | PR | F1 |
| W | **5070** | 303 | 165 | 9 | 102 | **89.8** | 84.8 | 87.2 | 84.4 | **91.1** | **87.6** |
| N1 | 657 | **2485** | 790 | 0 | 775 | 52.8 | **52.9** | 52.9 | **64.1** | 48.9 | **55.4** |
| N2 | 91 | 1113 | **27268** | 1204 | 389 | 90.7 | **92.3** | **91.5** | **91.5** | 91.2 | 91.3 |
| N3 | 7 | 2 | 696 | **6438** | 0 | **90.1** | **84.1** | **87.0** | 86.2 | 83.5 | 84.8 |
| REM | 153 | 798 | 628 | 0 | **9207** | **85.4** | 87.9 | 86.6 | 83.6 | **91.4** | **87.3** |

A visual comparison between the ground-truth stages and the stages predicted by our proposed method is shown in Fig.3.9 using the first subject"s recording from MASS. There are 1008 epochs available in total from this subject and the last 18 epochs are discarded so that the recording could be segmented into 30-epoch subsequences to feed the model.

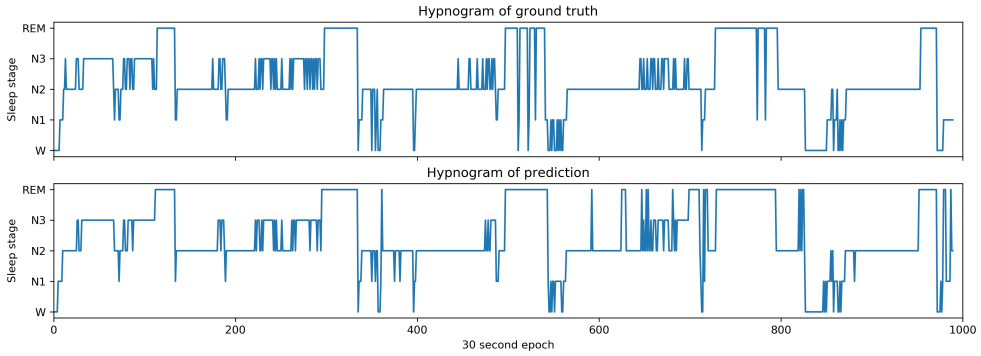


FIGURE 3.9: Hypnograms of subject 1 from MASS dataset. The top is ground truth and the bottom is prediction from our model. The accuracy and F1-score of our model are 0.87 and 0.82 respectively for this subject.

## 3.4 Ablation Studies

To validate the contribution of each component in our proposed architecture, we conduct comprehensive ablation studies on the Sleep-EDF dataset using the Fpz-Cz channel.

### 3.4.1 Component-wise Analysis

**Multi-scale Feature Extraction:** We evaluate the impact of our Hilbert transform-inspired multi-scale pooling by comparing against single-scale pooling. Results show that multi-scale pooling improves accuracy by 2.3% (from 82.0% to 84.3%), demonstrating the importance of capturing both frequency and amplitude features.

**Residual Blocks:** Replacing the two ResNext blocks with standard convolutional layers results in a 1.8% accuracy drop (84.3% to 82.5%), confirming the benefit of residual connections for gradient flow and representation learning.

**Self-attention vs. LSTM:** When replacing our multi-head self-attention with bidirectional LSTM while keeping other components unchanged, we observe: (1) 1.2% accuracy reduction (84.3% to 83.1%), (2) 8x increase in training time, and (3) inability to train end-to-end due to gradient vanishing issues.

**Cost-sensitive Learning:** Comparison between cost-sensitive learning and oversampling shows that while achieving similar accuracy (84.3% vs. 84.1%), cost-sensitive learning reduces training time by 65% and maintains better generalization on imbalanced test sets.

### 3.4.2 Architecture Depth Analysis

We experimented with different numbers of residual blocks (1, 2, 3, and 4 blocks). The results show that 2 blocks provide the optimal trade-off between performance and computational efficiency. Using only 1 block results in underfitting (accuracy drops to 81.7%), while 3 or 4 blocks lead to overfitting and increased computational cost without significant performance gains.

### 3.4.3 Attention Head Analysis

We varied the number of attention heads from 2 to 16. The optimal performance is achieved with 8 heads, consistent with the original Transformer paper. Fewer heads (2-4) result in insufficient representation diversity, while more heads (12-16) lead to overfitting and increased computational overhead.

**Justification for Limited Ablation:** Given the clear architectural inspirations (Hilbert transform for feature extraction, ResNet for representation learning, and Transformer for temporal modeling), and the significant computational savings achieved, we focus our ablation studies on the most critical components. The consistent improvements across two different datasets (Sleep-EDF and MASS) further validate the robustness of our design choices.

## 3.5 Conclusion

In this chapter, we present a fast end-to-end deep learning method for single EEG channel based sleep staging. Inspired by Hilbert transform which extracts meaningful features like amplitudes and frequency bands according to AASM, the architecture of the proposed network consists of three parts: feature extraction, encoder and decoder. During feature extraction, multi-scale max pooling layers are used to simulate Hilbert transform to extract features from raw signals. In encoder, two residual blocks are utilized to approximate analytic signals of input raw signals and output feature embeddings. A multi-head attention mechanism is incorporated in decoder part to capture temporal context across the embedding sequence and output the final prediction of stages.

Our comprehensive evaluation demonstrates both superior performance and computational efficiency. The proposed method achieves 84.3% accuracy on Sleep-EDF dataset, surpassing the previous state-of-the-art DeepSleepNet by 2.5%. More importantly, our method is 10-12 times faster in training and requires 40% less GPU memory, making it practical for clinical deployment. The ablation studies confirm that each component contributes meaningfully to the overall performance, with the multi-scale feature extraction providing the most significant improvement.

The computational cost analysis reveals that our efficiency gains stem from three key design choices: (1) elimination of two-stage training through residual connections, (2) parallel processing via self-attention instead of sequential LSTM computation, and (3) cost-sensitive learning avoiding data oversampling overhead. As there is no recurrence in our model and cost-sensitive learning is directly applied to handle the data imbalance issue instead of oversampling, the training of our proposed method can be sped up significantly while maintaining end-to-end trainability.

In our future research, we aim to improve the staging performance when dramatic transitions happen as shown in Fig.3.9 around epoch 550 and epoch 780, and explore the application of our efficient architecture to multi-channel EEG analysis.

CHAPTER 4

# Single-Channel EEG based Insomnia Detection with Domain Adaptation

---

In this chapter, due to the strong representation learning power of our model in the previous chapter, we utilize abundant EEG datasets to address the data scarcity issue in the detection of insomnia. We propose a domain adaptation based model to better capture insomnia related features of the target domain by leveraging stage annotations from the source domain. For each domain, two pairs of common encoder and private encoder are firstly trained to extract sleep related features and sleep irrelevant features, respectively. In order to further discriminate source domain and target domain, a domain classifier is introduced. Then, the common encoder of the target domain will be used together with the Long Short Term Memory (LSTM) network for insomnia detection. To the best of our knowledge, this is the first deep learning based domain adaptation model using single channel raw EEG signals to detect insomnia at subject level. We use the Montreal Archive of Sleep Studies (MASS) dataset which contains only healthy subjects as source domain and two datasets which contain both healthy and insomnia subjects as target domain to validate our model's generalizability. Experimental results on the two target domain datasets (a public one and an in-house one) demonstrate that our model generalizes well on two target domain datasets with different sampling rates. In particular, our proposed method is able to improve insomnia detection performance from 50.0% to 90.9% and 66.7% to 79.2% in terms of accuracy on the two target domain datasets, respectively.

## 4.1 Introduction

One third of the general population are suffering from insomnia [156] which is a potential cause of depression, diabetes, hypertension, cardiovascular disease [102]. The symptoms of insomnia include sleep difficulties or frequent awakenings in the middle of the night. The diagnosis of insomnia highly depends on recurring complaints from patients themselves at least three times a week for more than three months in the format of subjective and time-consuming sleep questionnaires [157]. Those complaints include difficulties to fall into sleep and consequent daytime repercussions like fatigue, unable to concentrate, anxious mood and other problems [158, 159].

To collect objective sleep data, polysomnograph (PSG) is the gold standard for most sleep related studies [160, 161] and consists of multivariate signals captured during sleep such as breathing, oxygen saturations, heart rate, electromyography (EMG) [162] and electroencephalogram (EEG) [4]. EEG is one of those signals collected via electrodes placed on the scalp of a subject to capture brain activities. A large body of evidence from quantitative EEG analysis has demonstrated spectral features of some wave bands can help understand each individual's sleep for the diagnosis of sleep diseases such as insomnia [163, 164, 165]. However, the first-night effect of PSG could significantly impact the subjects' sleep quality thus affect the diagnosis and identifying either sleep waves or sleep stages requires the interventions from highly trained sleep technicians, which makes the diagnosis of insomnia from PSG and sleep questionnaires a tedious and subjective process. Therefore, how to utilize computer algorithms for automatic and objective insomnia diagnosis has received great attention recently.

The traditional studies on EEG based insomnia detection focused on devising various hand-crafted features which provide reasonable explainability. Corsi-Cabrera *et al.* applied Fast Fourier Transform (FFT) on 2-second EEG epochs to extract Beta and Gamma wave bands from frontal and left hemisphere, then conducted statistic analysis according to temporal coupling of those two wave bands. It was noticed that insomnia patients had temporal alterations such as high Beta-Gamma and low Delta activities during their sleep [166]. Colombo *et al.* used amplitude envelope from band-filtered EEG oscillations to estimate Long-Range

Temporal Correlations (LRTC) of brain dynamics as LRTC positively associates with severity of insomnia [167].

Recently, with the rapid progress of machine learning, especially deep learning, in representation learning for various classification tasks such as object recognition and scene detection [168, 169, 170]. In medical research, many machine learning methods have also been proposed for EEG based insomnia detection. Dissanyaka *et al.* [171] pre-processed EEG data from 4 electrodes and extracted several frequency domain features from power spectrum estimation with Short-time Fourier Transform (STFT), Yule-Walker algorithm and Wavelet Transform. Those features were fed into a classification tree to differentiate insomnia patients from normal subjects. The results demonstrated that sleep onset period has significant contributions to the final prediction of sleep stages. On the other side, to evaluate the effectiveness of Singular Spectrum Analysis (SSA) in differentiating insomnia [172], a multi-layer feed forward Artificial Neural Network (ANN) was designed to diagnose insomnia with singular spectra computed from C3 and C4 EEG channels. The results showed that SSA can detect the oscillatory variations of sleep EEG and can be utilized to support the clinical findings for psycho-psychological insomnia disorders.

However, training a deep learning model from scratch for insomnia detection is a challenging task because of data scarcity in the field. To address the data scarcity problem, domain adaptation has been widely used to enhance representation learning for a target domain with limited training data by leveraging the data abundance of a source domain. For example, adults' data was used to diagnose infants' heart disease through domain adaptation [173]. On the other hand, although it is possible to filter out artifacts using signal processing techniques, it can also cause information loss. In addition, an effective model which uses single channel EEG signals can be easily generalized to multiple channels.

Therefore, based on the hypothesis that sleep stages have a close relationship with insomnia as mentioned in [167] and sleep or insomnia related features should be neutral from those factors which may impact diagnosis such as variance of subjects and equipment, in this work, we propose a novel domain adaptation based insomnia detection model to transfer features learned through sleep staging (source domain) to insomnia detection (target domain)

regardless of the data collection context using single channel raw EEG signals. The proposed model is composed of two key parts: pre-training and insomnia detection modeling. In the first part, sleep task specific features are learned through auto-encoders, task specific classifier, and domain adversarial training. For example, in the source domain of sleep staging, a private encoder extracts data specific (i.e., private features related to data acquisition context including devices, protocol and subjects) and task specific features (i.e., common features associated to sleep related wave bands) are extracted by a common encoder. Specifically, extracted common features which are neutral from the the data collection context are used for sleep staging to train a robust feature extractor. To enforce the learning of common features between source and target domains, a domain classifier is introduced with domain adversarial training. In the second part, insomnia related features (i.e., target domain specific common features) are fed into a long short term memory (LSTM) network for insomnia detection.

In summary, the main contributions of this work are as follows:

- We propose a novel domain adaptation based deep model to address the data scarcity issue of insomnia detection by leveraging abundant EEG data in the sleep staging domain. To the best of our knowledge, this is the first study to learn task irrelevant features (i.e., private features) and task specific features (i.e., common features) through domain adaptation for insomnia detection.
- We devise a domain classifier to enforce the learning of common features through domain adversarial training, together with domain task specific and auto-encoder based feature learning in the pre-training part.
- We conduct comprehensive experiments on both a public and an in-house insomnia datasets with different sampling rates to clearly demonstrate the effectiveness and generalizability of our proposed model.
- We adopt raw EEG signals to avoid information loss introduced by pre-processing so that the success rate of insomnia detection can be improved which is crucial to diagnose chronic diseases related to insomnia.

This chapter is organized as follows. Section II reviews the related works including rule based approaches for insomnia diagnosis and various machine learning based approaches. Section III introduces the details of our proposed method. Section IV presents comprehensive experimental results of our proposed model. Lastly, Section V concludes our study.

## 4.2 Related Work

In this section, related studies on EEG based insomnia detection are reviewed with two categories: rule based and machine learning based.

### 4.2.1 Rule based Insomnia Detection

In the early studies, various empirical rules were proposed by comparing EEG signals of healthy subjects and those with insomnia. These rules are generally based on sleep parameters or sleep micro-structures.

#### 4.2.1.1 Sleep parameters based rules

To provide a brief and valid assessment tools with sleep parameters, Morin *et al.* [174] proposed the Insomnia Severity Index (ISI) as a brief self-report instrument measuring a patient's perception of his or her insomnia. ISI is composed of seven items that evaluate the severity of sleep-onset, sleep maintenance, early morning awakening, satisfaction with current sleep pattern, interference with daily functioning, noticeability of impairment attributed to the sleep problem, and the level of distress caused by sleep problem. Each of these items is rated from 0 to 4 and based on the experience in the last 2 weeks. Total scores range from 0 to 28, where higher score means severer insomnia.

As ISI has been widely used in insomnia related sleep studies, Morin *et al.* [175] combined ISI with sleep diary to subjectively estimate sleep parameters including daytime nap, sleep aids intake, bedtime, sleep onset latency, frequency of nocturnal awakenings, awakenings duration, wake-up time, arising time, feeling upon arising and sleep quality, and showed ISI

is also adequate and sensitive to detect changes related to insomnia treatment. In the study, a cut-off score of 10 was used to determine whether an insomnia complaint was below the clinical threshold. Apart from that, Bastien *et al.* [176] conducted two studies with data from different samples of insomnia patients in order to validate the effectiveness of ISI. Results showed the ISI measurement has adequate internal consistency and is a reliable self-report measure to evaluate perceived sleep difficulties.

#### 4.2.1.2 Sleep micro-structure based rules

Cyclic Alternating Pattern (CAP), which characterizes sleep micro-structures, is defined as phasic arousals during Non-Rapid Eye Movement (NREM) sleep, consisting of repetitive spontaneous sequences of transient events (Phase A of sub-types A1, A2 and A3) with intervals of intermittent recovery (Phase B) that separate the repetitive elements [127]. Chouvarda *et al.* [177] hypothesized that sleep micro-structures characterized through CAP are different between normal and insomniac subjects and proposed to use normalized wavelet energy as CAP-related features of a single unipolar EEG derivation to develop insomnia detection rules. Insomnia patients have shorter phase duration and faster transition between phases while a normal subject's energy percentage is higher in a bigger scale than that of an insomnia subject. Furthermore, significantly higher complexity in short scales can be found in insomnia with respect to wavelet entropy. In [178], they further developed activation-deactivation duration based rules. An overall time in CAP sleep in NREM (CAPrate) is higher and the awake to sleep percentage is significantly higher while REM activity is lower in insomnia. With respect to the structure of A1 and B1 phases, wavelet energy has a maximum in scale 4, and a higher energy percentage in small scales for insomnia. With respect to CAP sleep, in A1 sub-type of CAP, a higher EEG complexity during B1 deactivation is related to insomnia.

### 4.2.2 Machine Learning Insomnia Detection

Machine learning based insomnia detection has become appealing recently due to its data-driven nature. In general, they can be categorized into conventional machine learning based approaches and deep learning based approaches.

#### 4.2.2.1 Conventional machine learning based approaches

Instead of designing detection rules empirically, machine learning based methods aim to learn the decision making from data through computational models.

Many conventional machine learning methods start with 6 sleep stages: wakefulness (Wake), non-rapid eye movement (Non-REM) stage 1 to stage 4 (S1-S4), and rapid eye movement (REM), which were first defined by Rechtschaffen and Kales (R&K) [50] in 1968. Then, in 2007, American Academy of Sleep Medicine (AASM) [51] revised those definition with 5 stages: Wake(W), Rapid Eye Movement (REM) sleep and Non-REM sleep stage 1 to stage 3 (N1-N3) by merging S3 and S4 from R&K rules into N3.

Conventional machine learning methods generally consist of two key steps: feature extraction and pattern learning. To imitate the decision making process of clinicians, Mulaffer *et al.* [105] used features from both EEG signals and hypnogram, and compared them with a support vector machine (SVM) classifier. A conclusion was drawn that hypnogram did have positive influence in modelling the insomnia diagnosis process. Recently, Sharma *et al.* [179] created eight different data subsets: individual stage (W, N1, N2, N3, REM) , LSS (light sleep stage, i.e., N1+N2) and ALL (all stages) from C4-A1 channel, then employed an optimal wavelet filters to extract norm features. Those features were fed into an ensemble bagged decision tree to classify insomnia epochs which achieved promising results.

Aiming to derive better features to understand sleep, Hamida *et al.* [180] investigated EEG time domain features like Hjorth parameters which quantify the mobility and complexity of EEG signals in terms of its variance and relative power of different wave bands. The results showed that primary insomnia subjects have lower theta power during wake and REM sleep while higher gamma power in wake, S1, S2 and REM stages and more beta activity during S1, S2 and REM. Then, in [181], they fed Hjorth parameters of pre-processed EEG signals from C3 and C4 channels into a K-means clustering based classifier and achieved 91% sensitivity in discriminating epochs from insomnia and control groups. Apart from that, they also applied PCA (Principal Component Analysis) to classify insomnia and control groups using 22 spectral features, i.e. ratios of relative powers and from different wave bands and

2 temporal features such as Hjorth parameters [182]. It achieved sensitivity of 91.6% using only deep sleep stage (SWS) epochs from EEG signals of C3 channel.

#### 4.2.2.2 Deep learning based approaches

There has been an increasing trend of applying deep learning to insomnia detection due to the resounding performance lately. With the assumption that sleep spindle plays a pivotal role in sleep continuity, Yu *et al.* [183] modelled the number of spindles and the chronological appearance of sleep spindles in a time window into sequences respectively, then applied an LSTM network to capture temporal patterns across the sequence to predict insomnia. The results demonstrated that temporal features of spindles instead of stationary features (i.e., frequency, duration, amplitude) are critical in building an automated insomnia diagnose system.

On the other hand, to investigate the correlation between insomnia and sleep stages, a staging model is developed in [104] using 57 features in time and frequency domains from 2 EEG channels in the first place and another model is created on top of it to discriminate insomnia patients. The experiment proved that insomnia-impacted stages are more dominant in insomnia prediction. The authors then proposed a two-stage algorithm in [184]. In the first stage, a deep learning based sleep staging model and an epoch-level insomnia detection model were obtained by feeding a set of temporal and spectral features derived from C3 and C4 EEG channels. The outputs of those two models were fed into a binary classifier to classify insomnia subjects.

In order to explore single channel EEG as a promising home-based approach for insomnia detection, Yang *et al.* [185] proposed a one-dimensional convolutional neural network (1D-CNN) model to provide epoch-level insomnia prediction. They derived 4 sub-datasets: All sleep stage dataset (ALL-DS), REM sleep stage dataset (REM-DS), light sleep stage dataset (LSS-DS), and SWS (slow wave sleep) sleep stage dataset from the EEG recordings of 9 insomnia patients and 9 healthy subjects, then conducted experiments under the intra-patient and inter-patient protocols. The results showed that REM and SWS epochs contribute most in

insomnia detection. However, none of the aforementioned approaches detects insomnia using subject level single channel raw EEG signals like our proposed method.

## 4.3 Proposed Method

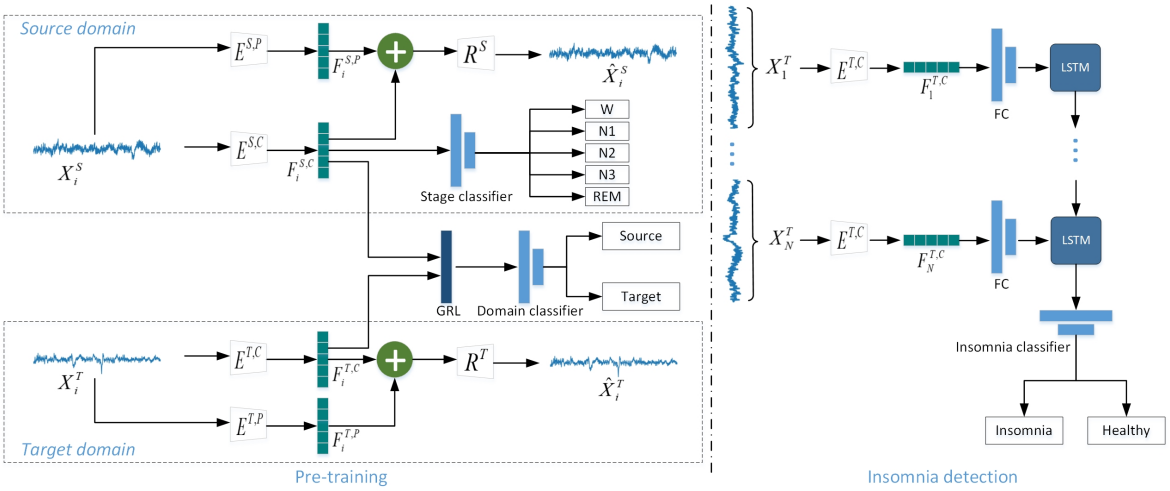


FIGURE 4.1: Overview of the proposed model which contains two parts: pre-training and insomnia detection. In pre-training, EEG epochs from source domain and target domain are fed into a pair of common encoders ($E^{S,C}$ and $E^{T,C}$) and a pair of private encoders ($E^{S,P}$ and $E^{T,P}$) respectively. A difference loss is applied between each pair of common and private encoders to make sure that the output from common encoders ($F^{S,C}$ and $F^{T,C}$) and private encoders ($F^{S,P}$ and $F^{T,P}$) are orthogonal. Meanwhile, a similarity loss is used to encourage $E^{S,C}$ and $E^{T,C}$ to extract common features from source and target domains, respectively. This is achieved by adding a Gradient Reversal Layer (GRL) before a domain classifier so that the classifier is confused and cannot identify which domain a sample is from. $R^S$ and $R^T$ make sure that the extracted features can reconstruct the original signals. In the insomnia detection phase, the sequence of encoded EEG epochs from $E^{T,C}$, which is supposed to contain information related to insomnia, is fed into an LSTM network to predict if a patient has insomnia or not.

As shown in Fig.4.1, our proposed method consists of two key parts: pre-training and insomnia detection. In pre-training, domain-invariant sleep related features are learned via a similarity loss between two common feature encoders from source and target domains. Domain-specific features are split from domain-invariant features by a difference loss and captured by a private feature encoder from each domain respectively. Meanwhile, a sleep stage classifier is trained

using the output of the common encoder in source domain to supervise the encoder to capture domain-independent patterns.

In the insomnia detection phase, the parameters in pre-trained model are frozen and the output of the common encoder in target domain is fed into an LSTM to capture insomnia-related temporal representations for final prediction.

### 4.3.1 Encoding-Decoding based Pre-training Module

Pre-training module is composed of feature extractors (encoders) and reconstructors (decoders) in each domain and a sleep stage classifier.

#### 4.3.1.1 Feature Extractor

A feature extractor $E$ aims to approximate the mapping between input $X$ and meaningful representations $F$ for a specific task, which can be mathematically defined as $F = E(X)$. Two feature extractors, $E^{S,C}$ and $E^{S,P}$, are used to extract common features $F^{S,C}$ and private features $F^{S,P}$ respectively from the sequences of 30-second EEG epochs in source domain $X^S$. Similarly, two feature extractors in target domain, $E^{T,C}$ and $E^{T,P}$, parameterize common representations $F^{T,S}$ across two domains along with private features $F^{T,P}$ from overnight EEG signals $X^T$ in target domain.

We took 600 consecutive EEG epochs from the middle of each recording into model to avoid the impact from wake stages and feed them as 20 sub-sequences with 30 epochs in each sub-sequence. To efficiently learn features from EEG signals, we adopt the architecture of our previous work on sleep staging [186] which has demonstrated strong representation learning capability in sleep staging task for all feature extractors. Each feature extractor consists of 3 modules: feature extraction, feature transformation and temporal context learning. The architecture of the feature extractor is described in Table 4.1.

The feature extraction module is formed by a convolutional layer with a filter size of $\lceil N/100 \rceil$ , where $N$ denotes the number of data points in each epoch (30 seconds * sampling rate) from

TABLE 4.1: Architecture of the feature extractors. It is composed by 3 modules: feature extraction, feature transformation and temporal context learning.

| module | architecture |
|---|---|
| feature extraction | fs conv, 64, /sd |
| | fs max pool, /sd \| [fs$*\pi$] max pool, /sd |
| feature transformation | 1 conv, 128, /2<br>3 conv, 64, /1<br>3 conv, 128, /1 |
| | 1 conv, 128, /1<br>3 conv, 64, /1<br>3 conv, 128, /1 |
| | global average pool |
| temporal context learning | positional encoding<br>multi-head attention |
| | 2048 fc<br>128 fc |
| | 5 softmax |

a sub-sequence. The stride is $\lceil N/1000 \rceil$ which is $1/10$ of the filter size. The output of the convolutional layer is fed into two different scale max-pooling layers. The pool sizes of those two max-pooling layers are $\lceil N/1000 \rceil$ and $\lceil \pi \times N/1000 \rceil$ respectively to simulate Hilbert Transform so that sleep related amplitude and time domain features can be properly captured. The strides of those pooling layers are set to the the same with the previous convolutional layer.

The feature transformation module aims to approximate the mapping between the output of extracted features from each epoch and the ground truth. To better extract features and avoid the gradient diminishing problem in very deep neural network, we refer to ResNext [143] which is one of the best performed variants of ResNet [142]. Two ResNext blocks [143] are stacked to transform extracted features into meaningful latent representations related to sleep. To prevent overfitting, a global average pooling layer [145] is used to average the feature maps obtained from residual blocks to produce embeddings for the next module.

In the temporal context learning module, a multi-head self-attention sub-module from Transformer [137] is utilized to learn temporal relationships across consecutive embeddings in a sub-sequence, because it is simply a feed-forward neural network, the module can be

efficiently trained. It linearly projects the latent temporal features into higher dimension using 8 heads and then bring the dimension of concatenated output from each head back to the same with input embedding using 2 fully connected layers for final prediction.

Two difference losses are introduced to push apart private and common representations in source domain and target domain. They are defined as the squared Frobenius norm of dot product between private and common feature vectors:

$$\mathcal{L}_{diff}^{S} = \| \ F^{S,C} F^{S,P} \ \|_F^2, \tag{4.1}$$

$$\mathcal{L}_{diff}^{T} = \| \ F^{T,C} F^{T,P} \ \|_F^2, \tag{4.2}$$

where $\|\|_F^2$ is the squared Frobenius norm. Minimizing different losses encourages orthogonality between common and private representations of each domain. In this work, they aim to separate those features in each domain.

#### 4.3.1.2 Domain Classifier

The domain classifier consists of two layers. The first layer is a Gradient Reversal Layer (GRL) which is the same with an identity layer, but reverses the gradients. The output of GRL is fed into the second layer, a fully connected layer, to predict which domain the input signal belongs to. GRL is used during training to confuse the classifier by reversing the calculated gradient (multiplying the calculated gradients by -1 in back propagation), which is adversarial to domain classifier's optimization. In this way, the model tends to make $F^{S,C}$ and $F^{T,C}$ more similar to increase the difficulty in differentiating which domain a sample is from. We exponentially change the multiplier of the gradients from 0 to -1 so that the model can be initialized properly in the beginning.

To ensure that the domain-invariant representations are effectively learned, domain-adversarial training [187] is introduced. This is achieved by utilizing a similarity loss between $F^{S,C}$ and $F^{T,C}$. The similarity loss is basically binomial cross-entropy loss. To handle data imbalance

from two domains, we decide to adopt cost-sensitive learning in training. Cost-sensitive learning assigns weights as penalty in loss function for those less accurately classified samples from minority classes. Thus, a weighted similarity loss function is defined as below:

$$\mathcal{L}_{sim} = w^D \sum_{i=1}^{N} \{d_i \log(\hat{d}_i) + (1 - d_i) \log(1 - \hat{d}_i)\}, \tag{4.3}$$

where $d_i$ is the ground truth domain label of the $i$-th sample while $\hat{d}_i$ is its predicted domain label and $N$ is the total number of samples from both domains. According to [186], $w^D$ denotes the contribution of a sample from a specific domain to the final loss. It can be derived from the inversed proportion of observations from a domain ($p^D$):

$$p^D = \frac{N^S + N^T}{N^D}, \tag{4.4}$$

where $N^S$ and $N^T$ stand for the number of observations from source domain and target domain, respectively, while $N^D$ is the number of observations from current domain.

Then, $w^D$ for a domain is defined below:

$$w^D = 2 \times \frac{p^D}{p^S + p^T}, \tag{4.5}$$

where $p^S$ and $p^T$ are the inverse proportion of observations from source domain and target domain. The more samples a domain has, the less contribution from the samples in this domain make to the final loss.

#### 4.3.1.3 Reconstructor

To minimize information loss of latent representations obtained from feature extractors and avoid trivial noise which could be highly correlated to classification [188], a reconstructor from each domain is introduced to ensure that input signals can be reconstructed by combining the learned representations from the private and common feature extractors.

The outputs of the common and private feature extractors are added before being fed to the reconstructor. The dimension of each feature vector is 128 and the feature map is resized using nearest neighbor interpolation to the original size of 30-second EEG echo which is $7680 \times 1 \times 1$ as the sampling rate in the source domain (MASS) is 256Hz and all signals from target domain will be resampled to 256Hz. Then, a convolutional layer with kernel size 7 and stride 1 is used to parameterize the mapping between the reconstructed and original signal. The output of this module is activated via a rectified linear unit (ReLU).

Instead of using traditional mean squared error loss which penalizes predictions that are correct up to a scaling term, scale-invariant mean squared error loss is applied to penalize differences between pairs of data points. The reconstruction loss functions for source domain and target domain can be defined as a sum of the scale-invariant mean squared error from target domain $T$ and source domain $S$ respectively as follows:

$$\mathcal{L}_{recon}^{S} = \sum_{i=1}^{N^S} \{ \frac{1}{K^S} \parallel X_i^S - \hat{X}_i^S \parallel_2^2 - \frac{1}{(K^S)^2} ([X_i^S - \hat{X}_i^S] \cdot 1_{K^S}) \}, \tag{4.6}$$

$$\mathcal{L}_{recon}^{T} = \sum_{j=1}^{N^T} \{ \frac{1}{K^T} \parallel X_j^T - \hat{X}_j^T \parallel_2^2 - \frac{1}{(K^T)^2} ([X_j^T - \hat{X}_j^T] \cdot 1_{K^T}) \}, \tag{4.7}$$

where $K$ is the number of data points in input $X$, $1_K$ is a vector of ones of length $K$ and $\| \cdot \|_2^2$ is the squared L2-norm.

#### 4.3.1.4 Stage Classifier

The stage classifier identifies the sleep stage of the $i$-th input EEG epoch in source domain $S$ by using the output $F_i^{S,C}$ of the common feature extractor $E^{S,C}$. The sleep stage classifier consists of a fully connected layer with 5 neurons.

The output $F_i^{S,C}$ of the common feature extractor $E^{S,C}$ is fed into the stage classifier as mini-batches of sub-sequences with a batch size of $O$ and a sequence length of $P$. The loss function is viewed as an average of the summed cross-entropy loss over all samples from source domain in the sub-sequence:

$$\mathcal{L}_{stg} = \frac{1}{O \times P} \sum_{i=1}^{O} \sum_{j=1}^{P} \mathcal{L}_{c,(i,j)}, \tag{4.8}$$

where $\mathcal{L}_{c,(i,j)}$ denotes the cross-entropy loss of the $j$-th sample in $i$-th sub-sequence, $O$ is set to 10 and $P$ is set to 30 in our work. As the prediction of an epoch depends on its neighbor, which is not related to the occurrence of a specific stage, cost sensitive learning is not applied here.

The cross-entropy loss of sleep staging is mathematically defined as:

$$\mathcal{L}_c = -\sum_{c=1}^{N_c} y_c^s \log(\hat{y}_c^s), \tag{4.9}$$

where $y_c^s$ is label of sleep stage $c$, $\hat{y}_c^s$ is the predicted probability of the observation in source domain and $N_c$ is the number of different sleep stages which is 5 in our work.

#### 4.3.1.5 Training Loss

The goal of pre-training is to minimize all losses mentioned above as follows:

$$\mathcal{L}_{pre} = \mathcal{L}_{stg} + \mathcal{L}_{diff} + \mathcal{L}_{sim} + \alpha \mathcal{L}_{recon} + \mathcal{L}_{reg}, \tag{4.10}$$

where $\mathcal{L}_{diff}$ is the sum of $\mathcal{L}_{diff}^S$(Eq. (4.1)) and $\mathcal{L}_{diff}^T$ (Eq. (4.2)), $\mathcal{L}_{recon}$ is the sum of $\mathcal{L}_{recon}^S$ (Eq. (4.6)) and $\mathcal{L}_{recon}^T$ (Eq. (4.7)), $\alpha$ is used to control the scale of reconstruction losses and is set to 0.01 for the best performance. In addition, L2 weight decay with parameter set to $10^{-3}$ is applied to the first convolutional layer of all feature extractors as additional regularization loss $\mathcal{L}_{reg}$.

### 4.3.2 Temporal Context based Insomnia Detection Module

After pre-training, the feature extractors can be used to obtain private and common representations. As common features are sleep task relevant and private features are relevant to data

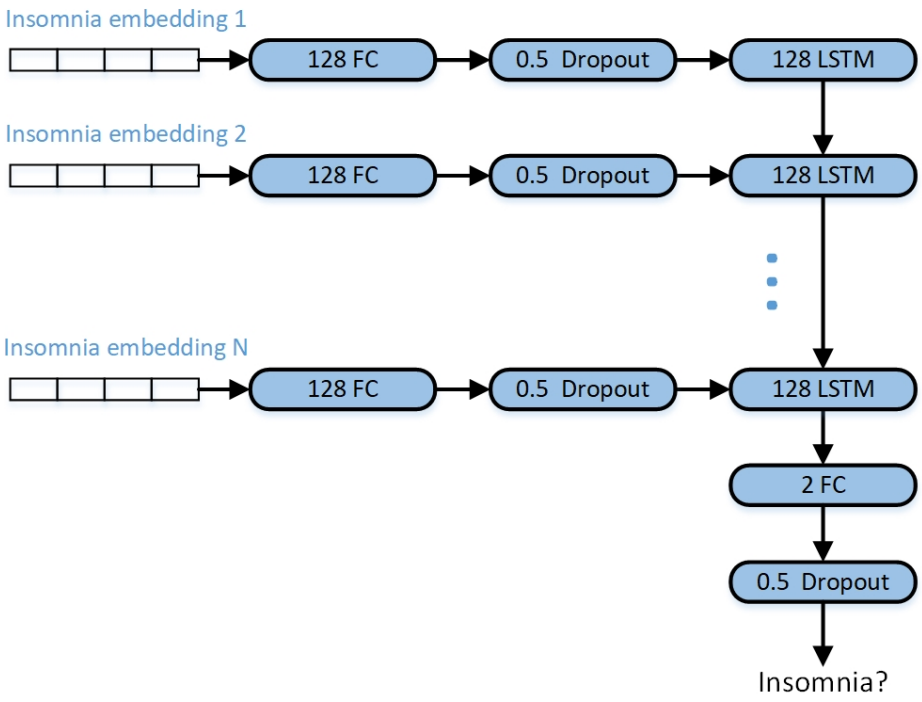


FIGURE 4.2: Illustration of the insomnia detection module. The embeddings from pre-training are fed into a fully connected layer followed by an LSTM layer and then another fully connected layer to classify insomnia subjects.

acquisition, the output of the common feature extractor in target domain is used to predict which overnight EEG signal is from an insomnia subject.

In the insomnia detection phase, we hypothesize that insomnia-invariant representations are pushed apart from insomnia-specific features and captured by common feature extractors. Therefore, we train an insomnia detection model on top of the output from $E^{T,C}$, which is supposed to learn insomnia related stage features in the fine-tuning stage. Cost-sensitive learning is also adopted to address data imbalance issue and the loss used in the fine-tuning stage is the weighted binomial cross-entropy loss:

$$\mathcal{L}_{ins} = w^M \sum_{i=1}^{N^T} \{m_i \log(\hat{m}_i) + (1 - m_i) \log(1 - \hat{m}_i)\}, \tag{4.11}$$

where $m_i$ is the ground truth insomnia annotation of the $i$-th sample while $\hat{m}_i$ is the prediction, $N^T$ is total number of recordings from target domain, and the calculation of $w^M$ is similar to Eq. (4.5).

The architecture of the insomnia detection module is illustrated in Fig.4.2. The extracted insomnia related feature embeddings of individual EEG epochs ($F_N^{T,C}$) are fed into a fully connected layer before an LSTM layer with 128 units and peephole connections [189]. We also explored other RNN variants like GRU (Gated Recurrent Unit) and Bi-directional LSTM while we found LSTM performed slightly better than GRU and Bi-directional LSTM was the worst. We also tried to add skip-connection for LSTM, however, it deteriorated the performance. To regularize the model, a weight decay with the degree of penalty $10^{-3}$ is applied to the fully connected layer prior to LSTM and a dropout layer with keep rate 0.5 is adopted to the output of the fully connected layer. To prevent LSTM from overfitting, a dropout with the probability of 0.5 is applied to all the units in LSTM. The output of LSTM is fed into a 2-unit dense layer followed by a dropout layer with 0.5 keep rate to produce final prediction.

## 4.4 Experimental Results and Discussions

In this section, we introduce the datasets used in our experiments, evaluation metrics, experimental settings, and performance analysis on insomnia detection with different domain adaptation settings.

### 4.4.1 Datasets

To evaluate the effectiveness and generalizability of our proposed model, we choose the popular Montreal Archive of Sleep Studies (MASS) dataset [122] for sleep staging as the source domain (only contains EEG signals from healthy subjects) and another two insomnia datasets as the target domains (contain EEG recordings from both healthy and insomnia subjects).

MASS contains whole-night recordings of EEG, EOG, EMG, ECG and respiratory signals from 97 males of ages from from 23 to 63 and 103 females of ages from 19 to 57. The recordings are in the European Data Format (EDF)++ format and organized into five subsets (named SS1-SS5) according to the research protocols. SS1 and SS3 were obtained via

montage formatting [190] annotated according to AASM while the others follow R&K rule. Since SS1 has a mixed montage size, we choose SS3 which is selected by plenty of sleep related work. The signals in SS3 were collected from 28 male and 34 female healthy subjects aged from 20 to 69 years old and contain 20 scalp-EEG channels, 2 EOG channels (left and right), 3 bipolar EMG channels (chin) and 1 ECG channel (D-I). The signals in this dataset were recorded at a mixed sampling rate of either 256Hz or 512Hz. To ensure the samples have the same dimension, the signals recorded at a sampling rate of 512Hz were downsampled to 256Hz uniformly.

One insomnia dataset is publicly available from Kermanshah University of Medical Sciences (Kermanshah) [191]. It consists of overnight (11PM-7AM) PSG recordings from 22 subjects which includes 8 males and 14 females aged between 18 and 63 years. 14 EEG channels (C4A1, C3A2, F3, F4, C3, C4, A1, A2, O1, O2, F3A2, F4A1, O1A2, O2A1), 6 EOG channels (EOG1, EOG2, EOG1A1, EOG2A1, EOG1A2, EOG2A2), 3 EMG channels (EMG, EMG1, EMG2) and ECG were collected at a sampling rate of 256 Hz. Half of those recordings are from insomnia patients while the other half are from healthy subjects. Those recordings also contain power spectrum features like alpha, beta, theta and delta waves which are annotated by sleep technicians according to AASM.

The other insomnia dataset is an in-house one collected at a sampling rate of 200Hz from 18 healthy subjects (6 males and 12 females aged between 18 and 50) and 30 subjects (11 males and 19 females aged between 18 and 80) who are diagnosed to have insomnia in Woolcock Institute of Medical Research (Woolcock) with the ethical approval from Royal Prince Alfred Hospital Ethics Review Committee, Sydney, Australia (Protocol No X11-0392 HREC/11/RPAH/620). Each recording contains 3 EEG channels (F4, C4, O2), 2 EOG channels (E1-M2, E2-M2), 2 EMG channels (EMG-L, EMG-R) and ECG. All subjects were requested to stay in the clinic for two consecutive nights. In the second night, they were restricted to the sleep time according to their subjective perception collected in the first night. This is one of the insomnia behavior treatments in order to measure whether the treatment has impacted a patient's brain signal during the sleep and enhanced their sleep quality. Since there

are no second night recordings from MASS, we only use the recordings from the first night in this work to eliminate factors introduced by human intervenes which may impact the result.

By following the same experiment protocol as described in [186], we select the F4 channel from both the source domain dataset (MASS) and the two target domain datasets (Kermanshah and Woolcock) to be consistent with most of sleep related research.

### 4.4.2 Evaluation Metrics

To evaluate sleep staging model on each class, we use precision (PR) to measure misclassification, recall (RE) to measure the model's capability in capturing patterns and F1-score (F1) to measure overall performance respectively. They are defined as follows:

$$PR = \frac{TP}{TP + FP}, \tag{4.12}$$

$$RE = \frac{TP}{TP + FN}, \tag{4.13}$$

$$F1 = 2 \cdot \frac{PR \times RE}{PR + RE}, \tag{4.14}$$

where $TP$, $FP$ and $FN$ are true positives, false positives and false negatives of a class respectively.

On the other hand, as both sleep staging and insomnia detection are classification tasks, two popular evaluation metrics are used to validate the overall performance: macro-averaging F1-score (MF1) and overall accuracy (ACC). MF1 and ACC are defined as follows:

$$ACC = \frac{\sum_{i=1}^{K} TP_i}{N}, \tag{4.15}$$

$$MF1 = \frac{\sum_{i=1}^{K} F1_i}{K}, \tag{4.16}$$

where $TP_i$ and $F1_i$ are the true positives and F1-score of the $i$-th class, respectively, $K$ is the total number of classes (i.e., 5 in sleep staging and 2 in insomnia detection), and $N$ is the total number of samples in the dataset.

### 4.4.3 Experimental Settings

To eliminate bias and ensure statistical significance, we adopt $k$-fold cross-validation to evaluate our model. Since there are only 22 subjects in the Kermanshah dataset, to maintain 80/20 ratio of train/test split, we set $k$ to 4. That means, an entire dataset is partitioned into 4 folds and the test split in each fold contains approximately same amount of healthy and insomnia subjects. Each subject is used only once in each fold and there is no overlap between training and test split. Similarly, to be consistent with Kermanshah dataset, 4 folds with train/test split ratio approximately 80/20 in each fold are created in the MASS and Woolcock datasets. The numbers of recordings in test splits for MASS, Kermanshah, and Woolcock in each fold are listed in Table 4.2. Since there is no validation split, no early stopping is applied. To make the input dimension on parity with MASS, we also upsample data by padding 28 zeros in the beginning and the end of each second from Woolcock segments.

TABLE 4.2: Number of **H**ealthy and **I**nsomnia subjects in the test split of each fold from the three datasets.

| Dataset | Fold 0 H+I | Fold 1 H+I | Fold 2 H+I | Fold 3 H+I | Total H+I |
|---|---|---|---|---|---|
| MASS | 16+0 | 16+0 | 15+0 | 15+0 | 62+0 |
| Kermanshah | 3+3 | 3+3 | 3+3 | 2+2 | 11+11 |
| Woolcock | 4+8 | 4+8 | 5+7 | 5+7 | 18+30 |

Since there is no existing research which classifies insomnia subjects using overnight single channel EEG recordings to the best of our knowledge, to evaluate the improvement introduced by our proposed domain adaptation method, we set the baseline by training a staging model using the same architecture with the feature extractor using stage annotations from source domain and fine-tuning the last layer with insomnia labels in target domain. During the fine-tuning step, we use the same architecture with that of the insomnia classifier described in Fig.4.2.

We adopt the Adam optimizer at a learning rate of $10^{-4}$ against the loss function defined in Eq. (4.10) during pre-training while in fine-tuning a Momentum optimizer with the same learning rate is used to minimize Eq. (4.11) for both our proposed model and the baseline. Our proposed method and the baseline were implemented with TensorFlow 1.12 [155] using Python 3.6.10 and the training was conducted on a computer with a GPU NVIDIA GeForce GTX 1080Ti. In pre-training, our proposed model converged in 150 training epochs while in insomnia detection, it converged in 50 training epochs. For each fold, it took approximately 3 hours and 20 minutes for pre-training and insomnia detection, respectively.

### 4.4.4 Hyperparameter tuning

We tried different batch size $O$ from 1, 5, 10, 30 and sequence length $P$ from 10 to 60 in Eq. (4.8) during model tuning. We also experimented 0.01, 0.05 and 0.1 for $\alpha$ in Eq. (4.10). To provide better regularization, we tested the model when weight decay was set to $10^{-3}$, $10^{-4}$ and $10^{-5}$ in the first convolutional layer of all feature extractors and the fully connected layer prior to the LSTM and $10^{-3}$ was chosen according to the results. In terms of the number of neurons of the fully connected layer and the number of cells in LSTM from the insomnia detection module, we selected 128 out of 8, 16, 32, 64, 256 and 512 eventually because the model achieved the best performance. Regarding the drop rate in all dropout layers and in LSTM cells, we noticed 0.5 rendered better result comparing to 0.2 and 0.8. In addition, learning rate of $10^{-3}$ was also tried out, but the model converged better when learning rate was set to $10^{-4}$.

### 4.4.5 Performance on MASS vs Kermanshah

The performance comparison on both sleep staging and insomnia detection between the above-mentioned baseline and our proposed model is shown in Table 4.3 where better performances are highlighted in bold. It is observed that while our proposed method's performance is comparable with the baseline on the sleep staging task, it clearly outperforms the baseline on the insomnia detection task with a significant margin. This clearly demonstrates the advantage of addressing data scarcity with domain adaptation.

TABLE 4.3: Overall performance comparison on sleep staging and insomnia detection using Kermanshah

| | Staging | | Insomnia | |
|---|---|---|---|---|
| | ACC | MF1 | ACC | MF1 |
| Baseline | **84.7** | **78.8** | 50.0 | 47.3 |
| Proposed method | 84.4 | 78.7 | **90.9** | **90.9** |

We further investigated the stage-wise classification performances of both the baseline and our proposed method as shown in Table 4.4 and Table 4.5, respectively. It is noticed that while both methods perform similarly overall, our proposed method performs better on W and REM in terms of F1-score and we assume this is due to that those stages are related to insomnia which are more likely discriminated by our method.

TABLE 4.4: Staging performance of the baseline in confusion matrix and classification metrics on MASS

| | Confusion Matrix | | | | | Classification Metrics | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **4562** | 496 | 75 | 9 | 195 | 83.5 | 85.5 | 84.5 |
| N1 | 386 | **2195** | 942 | 4 | 671 | 51.1 | 52.3 | 51.7 |
| N2 | 285 | 705 | **23297** | 944 | 784 | **91.1** | **89.6** | **90.3** |
| N3 | 6 | 0 | 839 | **5861** | 0 | 86.0 | 87.4 | 86.7 |
| REM | 222 | 903 | 428 | 3 | **7788** | 82.5 | 83.3 | 82.9 |

TABLE 4.5: Staging performance of our proposed method in confusion matrix and classification metrics using MASS vs Kermanshah adaptation

| | Confusion Matrix | | | | | Classification Metrics | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **4451** | 521 | 90 | 8 | 267 | 88.3 | 83.4 | 85.8 |
| N1 | 375 | **1844** | 1008 | 6 | 965 | 52.8 | 43.9 | 48.0 |
| N2 | 124 | 699 | **23195** | 859 | 1138 | **90.6** | 89.2 | **89.9** |
| N3 | 5 | 0 | 936 | **5751** | 14 | 86.8 | 85.8 | 86.3 |
| REM | 88 | 426 | 361 | 1 | **8468** | 78.0 | **90.6** | 83.9 |

The performance of insomnia detection in confusion matrix is shown in Table 4.6 for the Kermanshah dataset, which indicates that our proposed method significantly outperforms the baseline on all those three evaluation metrics. This clearly shows the effectiveness of our proposed method on leveraging the data abundance in sleep staging to address the data scarcity in insomnia detection through domain adaptation.

To better understand the model, we create the activation map of first insomnia patient's data from the Kermanshah dataset using the method proposed in [192]. We plot the raw data, activation heatmap and hypnogram of predicted stages in Fig. 4.3 from top to bottom. The $x$ axis represents epoch number and the $y$ axis represents voltage potential of EEG signal, heatmap amplitude and predicted stage, respectively. As some activations have less impact to final prediction, only those activations which are higher than 0.85 are visualized.

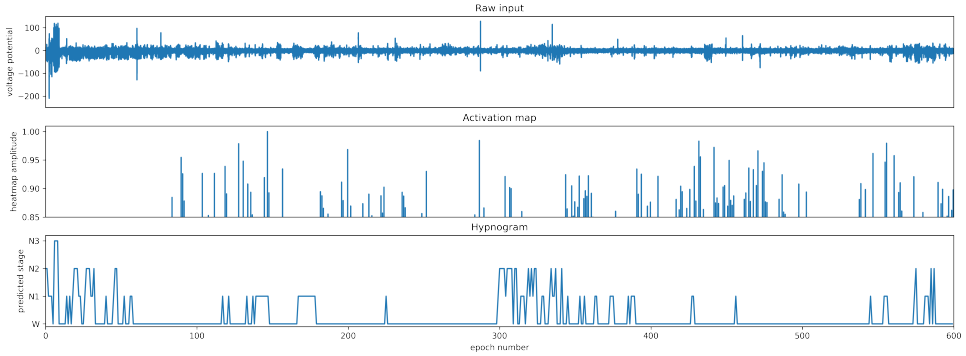


FIGURE 4.3: Visualization of the first insomnia patient in the Kermanshah dataset. The top figure is the raw signal in microvolts. As it contains significant artifacts, the normalized heatmap of activations higher than 0.85 with our insomnia model is visualized in the middle to locate key epochs. The bottom one is the hypnogram of predicted stages with our staging model.

From the hypnogram, we can see this subject almost stayed awake for the whole night. The highest activations are around the 146-th epoch which implies the difficulty of sleep onset. Around the 450-th epoch, the transition between W and N1 also contributes high activations to the final prediction of insomnia. Therefore, we can assume that this subject's sleep is very shallow and can be easily disturbed, which is consistent with what was described in [193].

To take a closer look at what happened in EEG, we plot EEG signal between the 145 and 149 epoch along with class activation map in Fig. 4.4. High frequency waves (mostly Beta-Gamma activities lower than 70Hz as mentioned in [166]) can be observed in those high activation windows, which demonstrates our model's effectiveness.

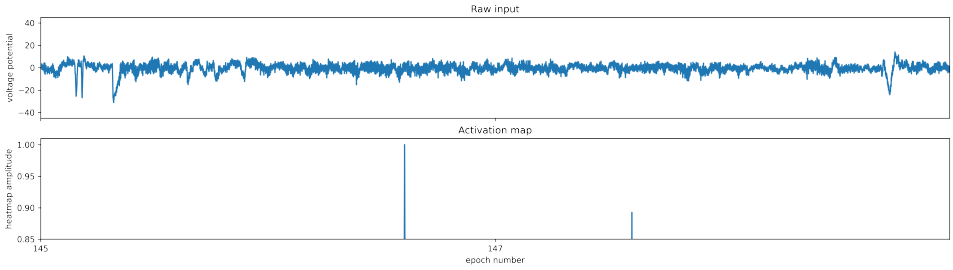


FIGURE 4.4: Raw signals (in microvolts) at the highest activation of the first insomnia patient in Kermanshah. Only those activations higher than 0.85 are visualized.

TABLE 4.6: Insomnia Detection Performance in Confusion Matrix and Classification Metrics using Kermanshah

| | Baseline | | | | | Proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | H | I | RE | PR | F1 | H | I | RE | PR | F1 |
| H | 8 | 3 | 50.0 | 72.7 | 59.3 | **10** | 1 | **90.9** | **90.9** | **90.9** |
| I | 8 | 3 | 50.0 | 27.3 | 35.3 | 1 | **10** | **90.9** | **90.9** | **90.9** |

### 4.4.6 Performance on MASS vs Woolcock

With respect to MASS vs Woolcock adaptation, the performance comparison on both sleep staging and insomnia detection between the baseline and our proposed model with MASS vs Woolcock adaptation is shown in Table 4.7 where better performances are highlighted in bold. Our proposed method outperforms the baseline on both the sleep staging task and the insomnia detection task, in particular with a clear performance improvement over the baseline on the insomnia detection task. Note that the overall insomnia detection performance on the Kermanshah dataset is better than that on the Woolcock dataset (90.9% vs 77.2%) even if Kermanshah dataset has fewer subjects (22 vs 48). We think it is because the filter size of the first convolutional layer in the feature extractor is calculated based on the MASS dataset (originally sampled at 256Hz) which can better capture patterns from the Kermanshah dataset (originally sampled at 256Hz) than the Woolcock dataset (originally sampled at 200Hz). However, it can demonstrate the effectiveness and generalizability of our method.

The detailed performance comparison of sleep staging for the baseline and the proposed method is shown in Table 4.8 and Table 4.9 with confusion matrix and classification metrics, respectively.

TABLE 4.7: Overall performance comparison on sleep staging and insomnia detection using Woolcock

| | Staging | | Insomnia | |
|---|---|---|---|---|
| | ACC | MF1 | ACC | MF1 |
| Baseline | 83.7 | 78.1 | 66.7 | 55.6 |
| Proposed method | **84.7** | **79.0** | **79.2** | **77.2** |

TABLE 4.8: Staging performance of the baseline in confusion matrix and classification metrics on MASS

| | Confusion Matrix | | | | | Classification Metrics | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **4600** | 515 | 57 | 9 | 156 | 76.6 | 86.2 | 81.1 |
| N1 | 480 | **2099** | 866 | 2 | 751 | 50.5 | 50.0 | 50.2 |
| N2 | 597 | 839 | **22682** | 872 | 1025 | **91.5** | **87.2** | **89.3** |
| N3 | 20 | 0 | 909 | **5768** | 9 | 86.7 | 86.0 | 86.4 |
| REM | 309 | 705 | 264 | 1 | **8065** | 80.6 | 86.3 | 83.4 |

TABLE 4.9: Staging performance of our proposed method in confusion matrix and classification metrics using MASS vs Woolcock adaptation

| | Confusion Matrix | | | | | Classification Metrics | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **4489** | 558 | 83 | 14 | 193 | 85.1 | 84.1 | 84.6 |
| N1 | 428 | **2040** | 990 | 7 | 733 | 50.0 | 48.6 | 49.3 |
| N2 | 202 | 828 | **23125** | 1168 | 692 | **91.2** | 88.9 | **90.0** |
| N3 | 5 | 1 | 691 | **6005** | 4 | 83.5 | **89.5** | 86.4 |
| REM | 149 | 656 | 479 | 1 | **8059** | 83.2 | 86.2 | 84.7 |

The detailed performance of insomnia detection is shown in Table 4.10. It is observed that our proposed method outperforms the baseline with a large margin in terms of Recall and F1-score. In particular, due to the imbalance of healthy and insomnia subjects, the baseline performs poorly in identifying healthy subjects, while this situation is significantly improved in our proposed method.

TABLE 4.10: Insomnia confusion matrix and per-class metrics from 4 fold cross-validation using Woolcock

| | | | Baseline | | | Proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | H | I | RE | PR | F1 | H | I | RE | PR | F1 |
| H | 4 | 14 | 66.7 | 22.2 | 33.3 | **12** | 6 | **75.0** | **66.7** | **70.6** |
| I | 2 | 28 | 66.7 | **93.3** | 77.8 | 4 | **26** | **81.3** | 86.7 | **83.9** |

On the other hand, similar to the activation map visualization of the first patient from Kermanshah, we also visualize the activations of the first insomnia subject 1 from Woolcock in Fig. 4.5 and Fig. 4.6 respectively. As per the hypnogram shown in Fig. 4.5, we can notice that this subject's sleep stages transit frequently. The highest activations are around the 36-th epoch. Around the 300-th epoch, the subject falls into deep sleep for the longest duration across the entire night. Dramatic changes among all sleep stages happen around the 320-th epoch, which shows the subject's awakenings during the night. Similar to the subject from Kermanshah, changes between W and N1 contribute to the very high activation around the 280-th epoch, which reflects difficulties in sleep maintenance. In the zoomed-in chart between the 34-th and the 38-th epochs from Fig. 4.6, wave bands with low amplitude and high frequency can be observed at those locations where activations are above 0.95.

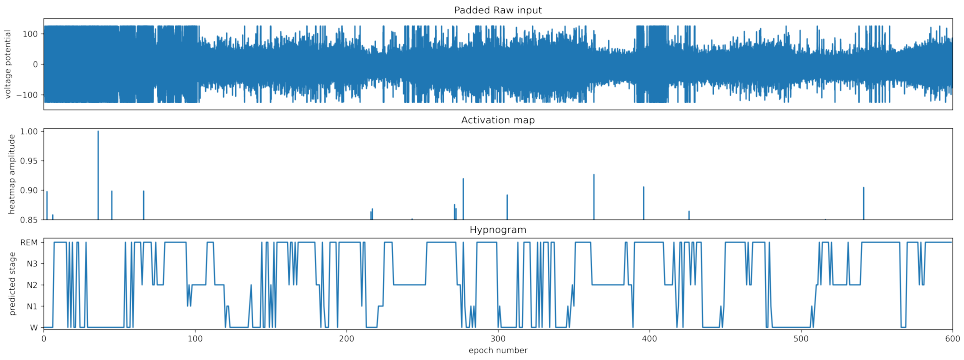


FIGURE 4.5: Visualization of the first insomnia patient in the Woolcock dataset. The top figure is the raw signal in microvolts. As it contains significant artifacts, the normalized heatmap of activations higher than 0.95 with our insomnia model is visualized in the middle to locate key epochs. The bottom one is the hypnogram of predicted stages with our staging model.

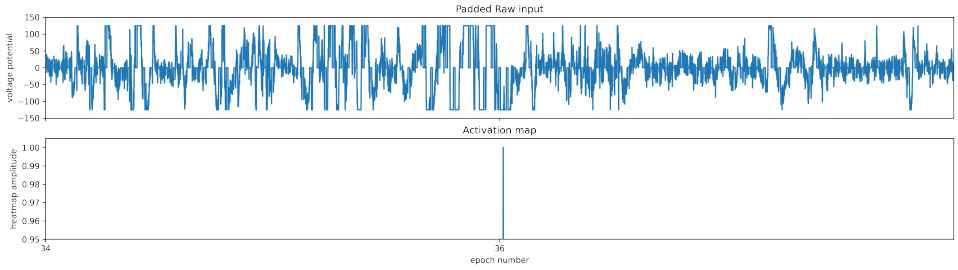


FIGURE 4.6: Raw signals (in microvolts) at highest activation of the first insomnia patient in Woolcock. Only those activations higher than 0.95 are visualized.

### 4.4.7 Computational Cost Analysis and Ablation Studies

#### 4.4.7.1 Computational Efficiency

To assess the practical applicability of our proposed method, we provide a comprehensive analysis of computational costs. Our model was implemented on a standard research workstation with NVIDIA GeForce GTX 1080Ti (11GB VRAM), Intel Core i7-8700K CPU, and 32GB RAM. The computational breakdown is as follows:

TABLE 4.11: Computational cost breakdown for different components

| Component | Training Time | Memory Usage | Parameters |
|---|---|---|---|
| Pre-training (150 epochs) | 3.3 hours | 8.2 GB | 2.1M |
| Insomnia Detection (50 epochs) | 20 minutes | 4.1 GB | 0.3M |
| Inference (per subject) | 0.8 seconds | 1.2 GB | - |
| Total Training | 3.6 hours | 8.2 GB | 2.4M |

The model's computational efficiency is justified by several factors: (1) The pre-training phase, while computationally intensive, only needs to be performed once for each source-target domain pair, making it a one-time cost; (2) The inference time of 0.8 seconds per overnight recording is clinically acceptable for routine diagnostic applications; (3) The relatively small model size (2.4M parameters) enables deployment on standard clinical hardware without requiring specialized high-end GPUs.

#### 4.4.7.2 Ablation Study

To validate the contribution of each component in our proposed architecture, we conducted systematic ablation studies by removing key components and evaluating their impact on insomnia detection performance. The results are presented in Table 4.12.

TABLE 4.12: Ablation study results on insomnia detection using Kermanshah dataset

| Model Variant | ACC (%) | MF1 (%) |
|---|---|---|
| Full Model | **90.9** | **90.9** |
| w/o Domain Classifier | 77.3 | 76.8 |
| w/o Private Encoders | 81.8 | 81.5 |
| w/o Reconstruction Loss | 86.4 | 86.1 |
| w/o Difference Loss | 83.6 | 83.2 |
| w/o Cost-Sensitive Learning | 86.4 | 85.7 |
| Baseline (No Domain Adaptation) | 50.0 | 47.3 |

The ablation study reveals several key insights:

- **Domain Classifier Impact**: Removing the domain classifier with GRL results in the most significant performance drop (13.6% in accuracy), confirming that adversarial domain adaptation is crucial for learning domain-invariant features.
- **Private Encoders Necessity**: The 9.1% performance reduction without private encoders demonstrates that explicit separation of domain-specific features is essential for effective knowledge transfer.
- **Reconstruction Loss Contribution**: The 4.5% drop without reconstruction loss validates that maintaining information integrity during feature extraction is important for preserving sleep-related patterns.
- **Difference Loss Significance**: The 7.3% performance decrease without difference loss confirms that enforcing orthogonality between common and private features improves feature disentanglement.
- **Cost-Sensitive Learning Value**: The 4.5% improvement with cost-sensitive learning demonstrates its effectiveness in handling class imbalance in medical datasets.

#### 4.4.7.3 Justification for Limited Ablation Scope

While extensive ablation studies are valuable, our focused approach is justified by several considerations specific to medical AI applications:

(1) **Clinical Relevance**: Our ablation study focuses on components that directly impact clinical performance rather than architectural variations that may not translate to practical diagnostic improvements.
(2) **Dataset Limitations**: With limited clinical data available (22 subjects in Kermanshah, 48 in Woolcock), extensive hyperparameter exploration could lead to overfitting to specific datasets rather than generalizable insights.
(3) **Computational Constraints**: In clinical settings, the emphasis is on deploying proven, stable architectures rather than extensively tuned models that may not generalize well to new patient populations.
(4) **Domain Adaptation Focus**: Since our primary contribution is the domain adaptation framework rather than novel architectural components, the ablation study concentrates on validating the effectiveness of adaptation mechanisms.

The computational efficiency and ablation results demonstrate that our proposed method achieves a favorable balance between performance and practical deployment considerations, making it suitable for real-world clinical applications while maintaining research rigor through systematic component validation.

## 4.5 Conclusion

In this chapter, we present a novel domain adaptation based deep learning method to detect insomnia using single EEG channel signals. It leverages the abundant data available in the sleep staging task to address the data scarcity issue in insomnia detection. The model consists of two key components: feature extraction and insomnia detection. The feature extraction component aims to extract domain specific features which are related to data acquisition and sleep task related common features which are associated to wave bands using two pairs of

encoders in source and target domain respectively. Two decoders are adopted to minimize information loss in each domain while one domain classifier is introduced to ensure sleep task specific features are extracted. Lastly, a sleep stage classifier is trained to approximate sleep stage related patterns and makes sure the input of domain classifier is sleep task related. The insomnia detection component feeds the sequence of extracted sleep-specific features from an overnight EEG recording by the target domain common encoder into an LSTM based binary classification model to predict whether a subject has insomnia or not. Experimental results on adapting the popular sleep staging dataset MASS (sampling rate 256Hz) to one public insomnia dataset (sampling rate 256Hz) and a larger in-house insomnia dataset (sampling rate 200Hz) clearly demonstrate the benefits of the proposed method for improving the performance of insomnia detection and generalizability on datasets with different sampling rates.

Our identified biomarkers, particularly the elevated Beta-Gamma activity patterns captured through activation maps, align well with established clinical understanding of insomnia pathophysiology. Clinical research has consistently demonstrated that primary insomnia patients exhibit specific waking EEG patterns characterized by hyperarousal states, including significantly higher beta and gamma power compared to healthy controls [166]. This hyperarousal manifests as increased activity in beta and gamma frequency bands associated with alertness and cognitive processing, along with reduced synchronization between frontal and posterior brain regions - patterns that persist even after sleep periods that would typically provide restorative effects in healthy individuals. The fact that our domain adaptation framework successfully identifies these clinically-validated Beta-Gamma biomarkers without manual feature engineering strengthens the clinical relevance of our approach and suggests that the extracted sleep-specific features capture meaningful neurophysiological signatures of insomnia. This convergence between our automated deep learning findings and established clinical EEG markers provides additional validation for the biological plausibility of our detected patterns and supports the potential clinical utility of our framework for objective insomnia assessment. In our future research, we aim to generalize our models to support multiple target domains and take subject-dependent characteristics into consideration for better performance.

CHAPTER 5

# CAPSleepNet: A Cyclic Alternating Pattern based Multi-view Sleep Staging Model

---

In this chapter, we combine the learning experience from previous studies to explore the possibilities to leverage finer-grained features within a 30-second epoch. We propose a novel method, called CAPSleepNet, which leverages the cyclic alternating pattern (CAP) events that represent the finer-grained temporal context within an epoch, such as sleep microstructure, to improve performance in sleep staging. However, CAP labels are not readily available in most staging datasets. To improve the accuracy of sleep staging, our method adopts a domain adaptation framework to adapt the epoch features to the CAP features so that they can be transferred from the source CAP dataset to the target staging dataset. Furthermore, to model various discrepancies between the source and target datasets, such as gender, age, sampling rate, etc., we view each sample as a fine-grained domain and turn the domain adaptation into a classification problem. Our method consists of five novel modules to facilitate the effective adaptation: (1) preliminary feature extraction, which extracts the preliminary feature vectors from 1-second EEG segments (2) epoch view extraction, which projects every 30 consecutive preliminary feature vectors into an epoch-level view (3) domain style extraction, which uses the epoch-level views as convolutional kernels to extract the instance-level domain styles (4) CAP view generation, which adapts the epoch-level view into the CAP views and reconstructs the missing target CAP pseudo-labels via a self-supervised approach (5) sleep stage classification, which combines the epoch-level view and the CAP view to classify sleep stages. The end-to-end training process is supervised by an inter-epoch Long Short Term Memory (LSTM) stage classifier, which takes both epoch and CAP features as input, and another intra-epoch LSTM CAP classifier simultaneously. We conducted experiments on two

public sleep staging datasets, i.e., Montreal Archive of Sleep Studies (MASS) and Sleep-EDF, and demonstrated that our proposed method outperforms the state-of-the-art.

## 5.1 Introduction

Understanding a person's sleep stages is essential for diagnosing sleep disorders. Sleep stages and rules were introduced by Rechtschaffen and Kales (R&K) [50] in 1968 to objectively measure sleep quality by scoring 30-second segmented polysomnography (PSG) signals, called epoch, into different sleep stages. R&K rules defined six sleep stages: wakefulness (Wake), four non-rapid eye movement (Non-REM) stages (i.e., Stage 1 to Stage 4), and rapid eye movement (REM). The American Academy of Sleep Medicine (AASM) revised and expanded the rules [51] in 2007. It defined three stages: wake (W), rapid eye movement (REM) sleep, and Non-REM sleep, where Non-REM sleep can be further split into Non-REM1 (N1), N2, and N3. In particular, N3 is formed by merging Stage 3 and Stage 4 of R&K rules.

The electroencephalogram (EEG), part of the PSG, captures brain waves via electrodes placed on the scalp of a brain and is frequently used as the gold standard for manual scoring of sleep studies. Sleep technicians score the sleep study using a combination of the EEG signal, EMG, EOG, and cardiorespiratory measured in 30 second epochs. The main signal used however is the EEG as it contains the different spectral waves which are the main feature of sleep stages differentiation. The process can be error-prone [53] and requires special domain knowledge, making it hard to scale. Therefore, research on automating clinical sleep staging has attracted attention nowadays.

Existing sleep staging research uses sleep macrostructure features extracted from sleep stages, neglecting temporary events such as K-complexes and transient power fluctuations in frequency bands [194, 195, 196]. The cyclic alternating pattern (CAP), which often gives critical information on sleep quality, especially in NREM sleep [5], is introduced to compensate for the limitations of the sleep macrostructure. CAP was established in 2001 as a

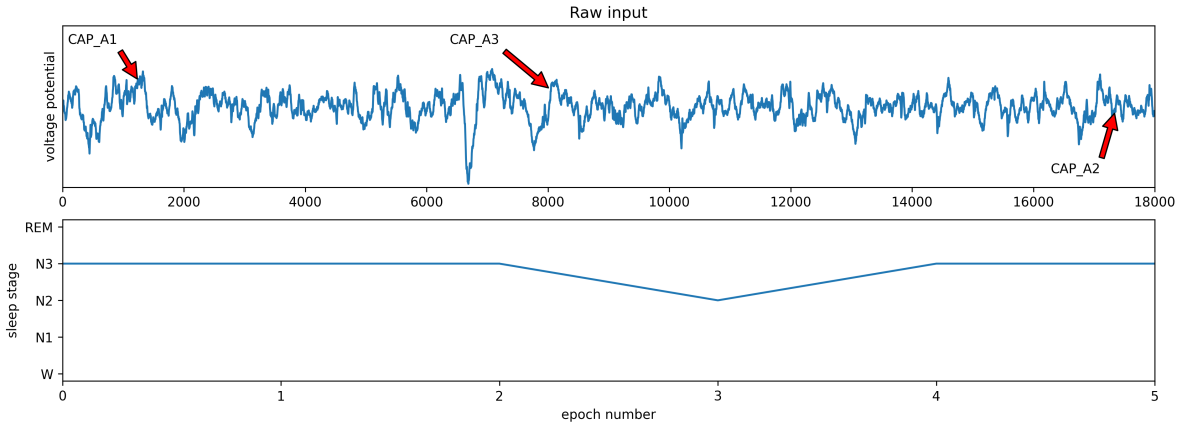


FIGURE 5.1: A sample of epochs which contain multiple CAP events. The A3 CAP event can be observed during the transition from N3 at epoch 2 to N2 at epoch 3.

physiological marker of sleep instability [197]. Thus, leveraging the CAP can benefit sleep staging tasks when discriminating hard samples, as shown in Fig. 5.1.

A CAP cycle consists of two phases, commencing with phase A, which is activation in the EEG signals, followed by periods of phase B, which is deactivation. Each phase duration can vary between 2 and 60 seconds [198]. Within phase A of CAP, there are three further sub-phases, namely A1, A2, and A3. Subtype A1 contributes to the build-up and consolidation of deep slow-wave sleep. Subtypes A2 and A3 contribute to the onset of REM sleep or wakefulness.

Due to the merit of CAP, it is finer-grained than a sleep stage and can help generate discriminative features for sleep staging. However, no research has been conducted on leveraging CAP in sleep staging to the best of our knowledge. To incorporate the CAP into sleep staging tasks, two challenges need to be addressed:

- Learned CAP features on the CAP dataset need to be transferred to the sleep staging datasets as the CAP annotations are not available in most sleep staging datasets.
- To ensure performance, multiple factors, such as the data collection protocols, the patients, the EEG channels, etc., between the source dataset and the target dataset need to be considered while transferring.

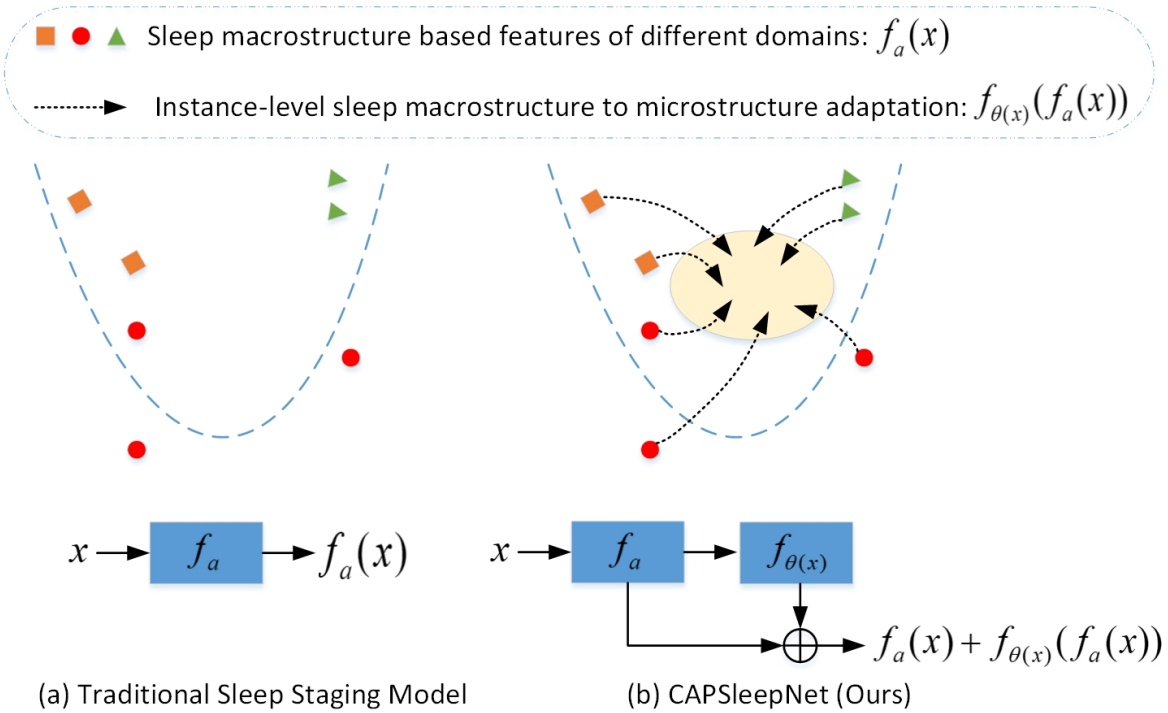


FIGURE 5.2: Traditional sleep staging model $f_a$ only extracts sleep macrostructure features. In CAPSleepNet, sleep macrostructure features are adapted into sleep microstructure features at the instance level by $f_\theta$ so that finer-grained discrepancies between the source dataset and the target dataset can have better class-wise alignment in classification. Both the sleep macrostructure and microstructure are considered in the model training.

This chapter combines the sleep macrostructure from the stage annotations with the sleep microstructure from the CAP annotations and feeds them into a multi-view deep learning staging model as shown in Fig. 5.2. To address the aforementioned challenges, we view each instance from the source dataset as a fine-grained domain and use a dynamic convolutional kernel to extract instance-level differences of various factors between the source dataset and the target dataset. To reconstruct the missing CAP labels on the target dataset, we first adapt the sleep stage-related features to the CAP features in the source dataset, then introduce a self-supervised approach to generate pseudo-CAP labels in the target dataset during the end-to-end transfer learning.

To the best of our knowledge, our method is the first in the sleep staging literature to utilize sleep microstructure in sleep staging. The result demonstrates its effectiveness compared to the state-of-the-art.

## 5.2 Related Work

### 5.2.1 Sleep Staging

In early research, signal processing techniques are frequently adopted to extract frequency and time domain features from raw EEG signals for sleep staging tasks. Typically, in these approaches, raw signals from EEG are fed into band-pass filters [58, 15, 23] to extract spectrum features for machine learning models. However, the signal processing approach requires domain knowledge of sleep signals for the preprocessing, and it can also cause information loss.

On the other hand, deep learning has gained significant attention in this field due to its strong learning power and capability to extract patterns from the raw data. DeepSleepNet [65] was proposed as the first deep learning approach using single-channel raw EEG signals for sleep staging. It uses two CNNs at different scales to capture time-domain and frequency-domain representations, respectively, and a bidirectional-LSTM (Bi-LSTM) was used against the concatenated output of those CNNs to capture temporal context among epochs.

Since then, plenty of research has been inspired to adopt deep learning. IITNet [199] captures inter- and intra-epoch temporal contexts from raw single-channel EEG. SleepEEGNet [200] feeds extracted time-invariant features into a sequence-to-sequence model to learn the relationship between the epochs and labels. SeqSleepNet [201] which consists of a filter bank layer to learn frequency-domain features and an attention-based recurrent layer for short-term temporal context followed by another recurrent layer to capture long-term inter-epoch features, and our previous work [186] which models the Hilbert transform to extract features and feeds them into a transformer for temporal learning. Those approaches have demonstrated the success of deep learning in the sleep staging field. Among them, TinySleepNet [202], which

significantly reduces parameters by removing one CNN from DeepSleepNet and replacing Bi-LSTM with an LSTM, achieves state-of-the-art and generalizes well on multiple EEG datasets.

To further improve the performance, recent research have proved that incorporating another view of EEG signals can improve the model's performance in sleep staging tasks. Ye *et al.* argues that the spectrogram can be complementary in identifying hard samples [203]. Thus, the spectrogram is added as another view to cross-validate the similarities between learned representations from the EEG signals and the spectrograms. This way, the semantically similar positives from the two views are obtained. In the meantime, the negative candidates are enlarged to improve the model performance. Phan *et al.* identifies the multi-view models often underperform comparing the best single-view counterparts due to simple strategies like concatenation [204]. To address this, time-frequency images are utilized as another view for the model, and learning on the overfitted view will be discouraged. The results outperform previous state-of-the-art results. In [205], authors challenge that most of the existing multi-view sleep staging models only consider time series and images. While those inputs are in Euclidean space, features in non-Euclidean space, such as graphs, are also important. With that hypothesis, they apply the short-time Fourier transform to extract Euclidean space-related features from multiple sleep signals from different parts of the human body, e.g., brain, heart, eyes, and lower jaw, to construct one view of the sleep staging model. Another view is formed by the topological information of the human body, where those signals are learned by adaptive graph learning. Competitive results are achieved in this approach.

However, sleep microstructure has not been considered in the existing research so far.

### 5.2.2 Domain Adaptation

Domain adaptation aims to adapt a model from a labeled source domain to an unlabeled target domain. Most domain adaptation methods seek to minimize the distribution discrepancy between these two domains using distance metrics, such as Maximum Mean Discrepancy (MMD) [206], optimal transport [207], or Kullback-Leibler (KL) divergence [208]. In contrast,

others exploit adversarial training [209]. However, it is more difficult for multi-source domain adaptation (MSDA) where multiple factors need to be considered between the source dataset and the target dataset.

Most MSDA methods are also based on distribution alignment using a static model. Peng *et al.* collected a large MSDA dataset containing six domains and proposed a new deep learning approach to transfer knowledge from multiple labeled source domains [210]. This is achieved by aligning the distributions from the source domains with the target domain and simultaneously aligning the source domain instances with each other. The first empirical research demonstrates that aligning the source domain benefits MSDA tasks. Wang *et al.* constructed a knowledge graph under the guidance of correlated prototypes of various domains [211]. A Relation Alignment Loss is designed to exploit the relational interdependency among categories during domain adaptation so that the feature distributions can be better aligned across the source and the target domain. In [212], to effectively address the MSDA problem, the Curriculum Manager is introduced as an adversarial agent to update the curriculum during training dynamically and iteratively. It learns which domains or samples are best suited for the alignment to the target so that the latent domains with different transferability to the target distribution can gradually converge according to various levels of importance.

However, when the target domain is too coarse-grained in MSDA setting, a static model may fail to align all the domains completely. Therefore, Dynamic Instance Domain Adaptation (DIDA) [213] is proposed to view each instance as a fine domain, assuming that the cross-domain feature distribution consistency can be invalid. The traditional feature alignment across domains is no longer required by aligning classes at the instance level. Thus, the domain adaptation problem is turned into a classification problem as each fine domain is at the instance level. A dynamic neural network with multi-scale adaptive convolutional kernels is developed to produce instance-adaptive residuals to model domain style into deep features for each instance. A semi-supervised learning approach is also applied to generate pseudo labels for the unlabeled target data using a cross-entropy loss. In the end, the evaluation metrics and experiment results are discussed.

## 5.3 Proposed Method

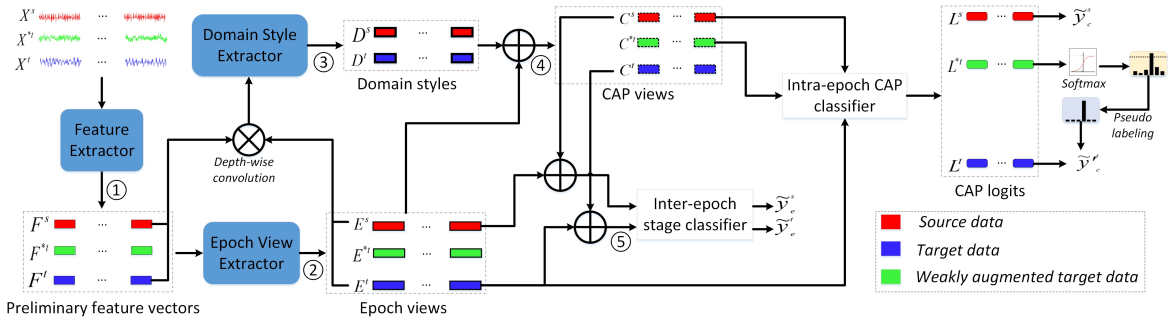


FIGURE 5.3: Our proposed model comprises five modules. 1. Preliminary feature extraction: input sequences of EEG epochs from the source and the target datasets, as well as a weakly augmented version of target signals, are fed into a feature extractor in batches. The feature extractor projects each 1-second input segment into a preliminary feature vector. 2. Epoch view extraction: the epoch view extractor takes 30 consecutive feature vectors and projects them into higher dimensions as epoch-level views. 3. Domain style extraction: the epoch-level views are used as dynamic convolutional kernels to extract the domain styles across both the source and target datasets at the instance level from the preliminary feature vectors. 4. CAP view generation: the CAP views from the source and augmented target signals are obtained from the epoch-level views by adding the corresponding domain styles. The missing CAP view from the target dataset is reconstructed during the CAP classification by feeding the generated source CAP view, augmented target CAP view, and the target epoch view into an intra-epoch CAP LSTM classifier. The CAP logits are produced for source, target, and weakly augmented signals, respectively. High-confidence predictions from weakly augmented target domain signals are used as pseudo labels to supervise the training. 5. Sleep stage classification: the extracted CAP views are combined with the epoch views from both the source and the target datasets as the input of an inter-epoch LSTM stage classifier.

This work considers our problem as a multi-view MSDA problem on 1-D EEG signals. Therefore, we are inspired by DIDA and generalize it to a multi-view model illustrated in Fig. 5.3. It comprises five modules: preliminary feature extraction, epoch view extraction, domain style extraction, CAP view generation, and sleep stage classification.

### 5.3.1 Preliminary Feature Extraction from 1-second EEG Segments

With the assumption that discriminative representations of the CAP features and the sleep stage features can be retrieved from 1-second EEG splits and 30-second EEG epochs, respectively, this module takes input batches of subsequences and extracts preliminary features for the downstream modules. An input batch contains $N$ sequences, each containing $L$ consecutive epochs. $N$ sequences are formed up by $N^s$ sequences from the source dataset ($X^s$), and $N^t$ sequences from the target dataset ($X^t$) as well as $N^t$ weakly augmented sequences from the target dataset ($X^{*t}$). In other words, $N$ equals $N^s + N^t * 2$.

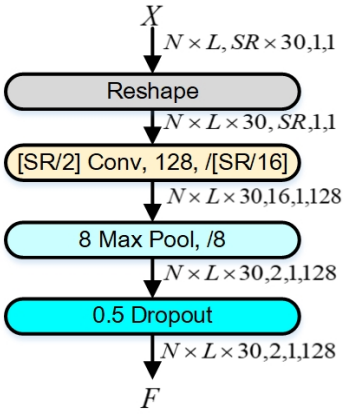


FIGURE 5.4: Preliminary feature extractor. SR is the sampling rate. It takes 1-second EEG segments as input $X$ and projects them into the preliminary feature vectors $F$.

The design of the feature extraction module is as demonstrated in Fig. 5.4. The input batches are reshaped into new batches of 1-second EEG segments with a size of $N*L*30$. In a particular batch input $X$, a convolutional filter with the size of half of the sampling rate of EEG (SR) is used to extract meaningful features from the raw signals of every second. Meanwhile, the stride is set to $1/4$ of the sampling rate.

The output of the first layer is down-sampled by a max pooling layer with the pool size and the stride equal to 8. A dropout layer with a drop rate of 0.5 is used after the max pooling layer to provide regularization. Feature vectors $F$ will be generated for the next module to use.

### 5.3.2 Epoch View Extraction from 30-second EEG Segments

This module takes every 30 consecutive feature vectors from the preliminary feature vectors for the epoch view extractor to project the preliminary features into higher dimensions. The architecture of the epoch view extractor is as shown in Fig 5.5.

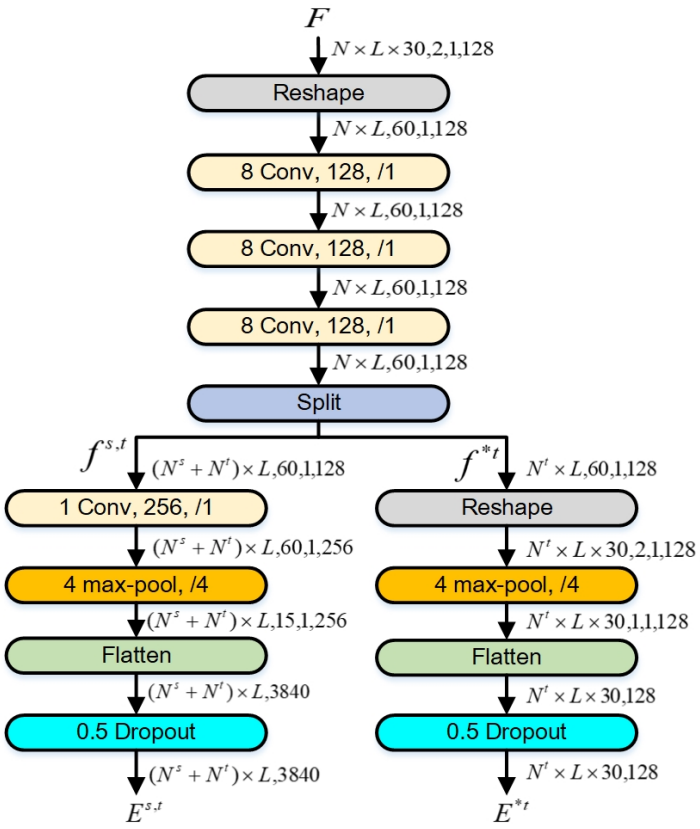


FIGURE 5.5: Epoch view extractor. It takes 30 consecutive feature vectors from the previous output $F$ as an epoch-level input and produces the intermediate feature vectors ($f$). The source and target domain intermediate feature vectors ($f^{s,t}$) are projected into higher dimensions by a convolutional layer with 256 kernels and fed into the downstream layers similar to the weakly augmented target domain intermediate feature vectors ($f^{*t}$) to generate the epoch views $E^{s,t}$ and $E^{*t}$.

Batches of the preliminary feature vectors are reshaped back to $N * L$, so the epoch view extractor can process at the epoch level. Three convolutional layers are incorporated to project the representations into higher dimensional space. Each convolutional layer has 128 convolutional kernels with a size of 8 and a stride of 1.

After being processed by the three convolutional layers, the feature vectors $f^{*t}$ from the weakly augmented target domain signals are split from $f^{s,t}$, the representations for the source and the target domains. Then, the channels of $f^{s,t}$ are increased to 256 using a $1 \times 1$ convolutional layer to facilitate later downstream addition. To further reduce dimension and extract meaningful features, those two extracted feature vectors are fed into a max pooling layer of size 4 and stride 4, followed by a flattened layer, respectively, to generate $E^{s,t}$ for the later stage classification and $E^{*t}$ for the target domain CAP pseudo labeling process.

### 5.3.3 Multi-Source Domain Style Extraction

To model the domain styles across different domains, we assume that sleep stages and CAP events share some representations [127]. Thus, we design a domain-style extraction module to generate domain-related patterns that can adapt the features learned from the instance-level domains to each other.

Inspired by DIDA, we view each instance as a fine domain, which requires the parameters of convolutional kernels to be conditioned on each instance as per the prior research [214], [215]. To explain our model architecture, we must understand Multi-Source Unsupervised Domain Adaptation (MSUDA).

To adapt a model trained on $k$ source domains, $\{S_1, ..., S_k\}$ to a target domain $T$ where $k > 1$, if each source domain contains labeled training samples $S_k = \{(x_i^{S_k}, y_i^{S_k})\}_{i=1}^{N_{S_k}}$ with $N_S$, $x$ and $y$ denoting number of source samples, data and label respectively, it can be viewed as a multi-source domain adaptation (MSDA). In particular, with the assumption that the target domain shares the same label space with the source domains and all its training data are not labeled represented as $T = \{x_i^T\}_{i=1}^{N_T}$, it is defined as Multi-Source Unsupervised Domain Adaptation (MSUDA).

In this work, sleep stages are annotated in both the source dataset and the target dataset, but CAP annotations are only available in the source dataset on a different EEG channel from the target datasets. Therefore, we view it as an MSUDA problem, and the focus of this chapter

is to improve the classification of sleep stages on the target datasets by adapting the CAP patterns from the source dataset.

We design the domain style extractor as shown in Fig 5.6. We adopt dynamically generated convolutional kernels to implement the instance-conditioned sub-module to tackle the scale variation between the epoch and CAP views. There are two branches in this module. The left branch reduces the channels of input $F^{s,t}$ to 16 denoted as $\mathcal{F}_{\theta_1}(F^{s,t})$ where $\mathcal{F}$ is the transformation done by the convolutional layer and $\theta_1$ denotes the parameters. The right branch is the kernel generator branch. As the input $E^{s,t}$ is originated from $F^{s,t}$, we have

$$E^{s,t} = \mathcal{T}(F^{s,t}), \tag{5.1}$$

where $\mathcal{T}$ represents the transformation from the epoch view extractor, objects are captured at different scales in this way. Thus, we reduce $E^{s,t}$ to 16 channels representation $\mathcal{F}_{\theta_2}(E^{s,t})$ and transpose the dimensions of the output so that it can be used as convolutional kernels and fused with $\mathcal{F}_{\theta_2}(E^{s,t})$ via depth-wise convolution:

$$O^{s,t} = \mathcal{F}_{\theta_2}(E^{s,t}) \otimes \mathcal{F}_{\theta_1}(F^{s,t}), \tag{5.2}$$

where symbol $\otimes$ refers to depth-wise convolution and the $O^{s,t}$ are the fused features. According to Eqn 5.1, above can be rephrased to:

$$O^{s,t} = \mathcal{F}_{\theta_2}(\mathcal{T}(F^{s,t})) \otimes \mathcal{F}_{\theta_1}(F^{s,t}), \tag{5.3}$$

The two fused feature vectors depend on the preliminary feature vector $F^{s,t}$, representing the original input instance $X^{s,t}$, contributing to instance adaptation.

A convolutional layer of 128 $1 \times 1$ kernels brings the channels of $O^{s,t}$ back to 128. Then, a dropout layer with a keep rate of 0.5 is appended, followed by a max pooling layer with pool

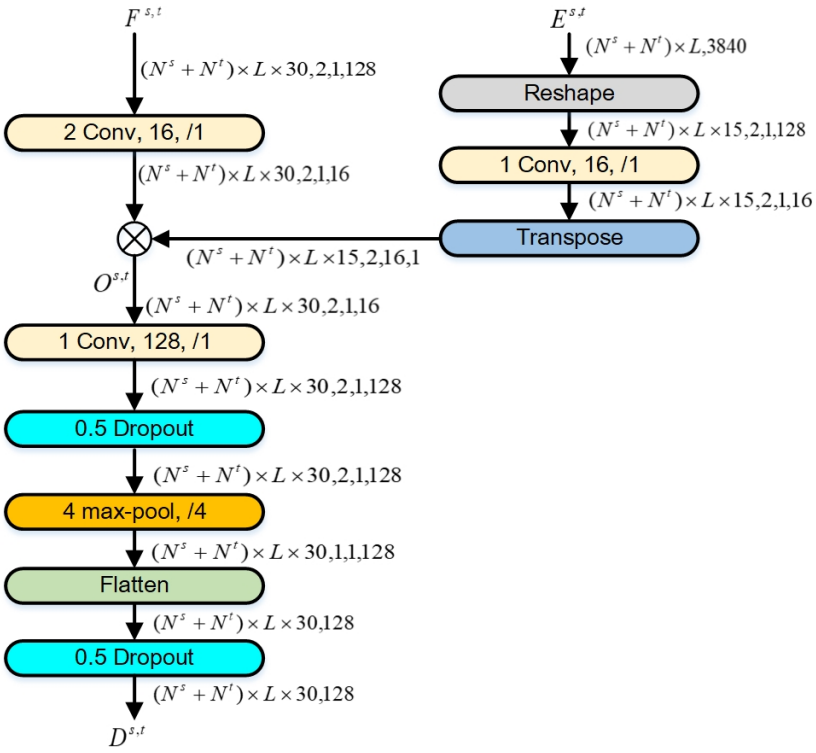


FIGURE 5.6: Domain style extractor. It takes preliminary feature vectors from the source and target domains ($F^{s,t}$). It uses the corresponding epoch view $E^{s,t}$ as a dynamical convolutional kernel to produce domain-related patterns $D^{s,t}$ via depth-wise convolution.

size and stride equal to 4. A flatten layer and another dropout layer with a 0.5 keep rate are applied to generate the output of this module $D^{s,t}$, which is deemed the domain style.

### 5.3.4 CAP View Generation and Pseudo Labeling

This module generates the CAP views from the epoch-level views for both the source and target datasets. Due to the lack of CAP annotations on staging datasets, pseudo labels will be constructed to generate the CAP views on staging datasets. A CAP classifier supervises the process.

Since we want to associate CAP events with corresponding sleep stages and CAP events only appear in NREM stages, we define six classes for CAP classification: A1, A2, A3, Unknown Wake (UW), Unknown REM (UR) and Unknown (U), where the first 3 classes are from the

raw labels and the last 3 classes are from Wake, REM and unknown stages, respectively. This is to ensure the CAP labels are available in all stages.

To effectively learn the temporal information from the CAP events, we adopt an LSTM on top of the CAP view generated by adding domain styles to the epoch view. The design of the CAP classifier can be depicted as Fig 5.7.

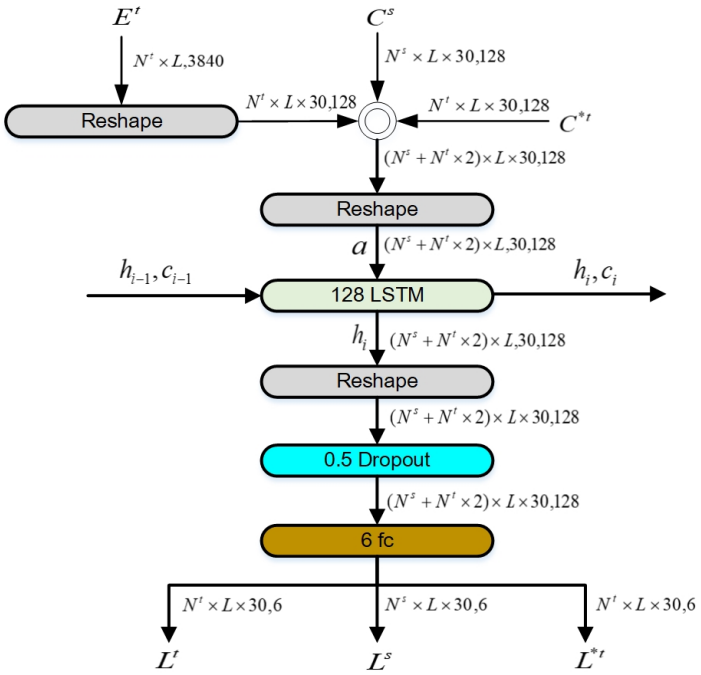


FIGURE 5.7: Design of the CAP classifier. All the input from upstream modules is concatenated in a batch and fed into an LSTM followed by a fully connected layer to generate the final logits accordingly.

The feature vector $a$ comprises the CAP view from the source domain $C^s$, the epoch view from the target domain $E^t$, and the CAP view of the augmented target domain signals $C^{*t}$. The CAP views can be obtained by adding the domain style to the epoch views, which means $C^s$ can be defined as:

$$C^s = E^s + D^s, \tag{5.4}$$

and $C^{*t}$ can be depicted as:

$$C^{*t} = E^{*t} + D^t, \tag{5.5}$$

The LSTM processes the features $a$ as follows:

$$h_i, c_i = LSTM_\theta(h_{i-1}, c_{i-1}, a), \tag{5.6}$$

where $\theta$ represents the learnable parameters of the LSTM, $h_i$ and $c_i$ are vectors of hidden and cell states of the LSTM layer after processing $a$. The hidden state is used for the final prediction by applying a softmax layer on top of it. Before feeding the hidden state into the softmax layer, a dropout layer with a keep rate of 0.5 is added to regularize the network. The output of the softmax layer contains logits formed up by $L^t$, $L^s$ and $L^{*t}$, which are derived from $E^t$, $C^s$, and $C^t$ respectively.

Since the CAP labels are unavailable in the target domain, we introduce a new method to generate the pseudo CAP labels inspired by FixMatch [216]. FixMatch is an algorithm that significantly simplifies the semi-supervised learning (SSL) methods that leverage unlabeled data to improve a model's performance. For a given image, FixMatch first generates pseudo-labels using the model's predictions on weakly augmented unlabeled images, and the pseudo-label is only retained if the model produces a high-confidence prediction. Then, the model uses the pseudo label to supervise the training with the strongly augmented version of the same image. The model demonstrated improved performance compared to the state-of-the-art due to its simplicity.

In our scenario, we apply random flip and crop as a weak augmentation on the target domain 1-second EEG splits with a probability of 50% to obtain $X^{*t}$. Then, it is passed through into the feature extractor along with the raw signals $X^s$ and $X^t$. After that, the feature vectors are fed into the epoch view extractor and added to the domain style to form the CAP view $C^{*t}$.

On the other hand, instead of applying the strong augmentations, which are hard to 1-D signals, we use the epoch view derived from the original target domain signals $E^t$ because we assume the epoch view is a strong augmentation from the CAP view.

We use two loss functions to supervise training the labeled source domain CAP and unlabeled target domain data, respectively.

For those labeled source domain data, we leverage cost-sensitive learning as we did in our prior work [186] to the class-weighted cross-entropy loss function:

$$\mathcal{H}(p, p') = -w_i \sum_{i=1}^{C} p_i \log(p'_i), \tag{5.7}$$

where $p_i$ is the actual probability an observation belongs to class $i$ and $p'_i$ is the predicted observation probability. $w_i$ denotes the weight of each class's contribution to the final loss, and $C$ is the number of classes that need to be classified.

Therefore, the loss function for the learning on the source domain labeled CAP data is:

$$\mathcal{L}_c^s = \frac{1}{M} \sum_{i=1}^{M} \mathcal{H}(p_i, p_c(\widetilde{y}_i^s | X_i^s)), \tag{5.8}$$

where $M$ is the number of the labeled CAP samples in a batch and $p_c(\widetilde{y}_i^s | X_i^s)$ is the predicted CAP label distribution produced by the CAP view $C_i^s$. Class weights 1.5 and 0.2 are assigned to A2 and U, respectively. The weights of the other four classes are set to 1.

Regarding the unlabeled data, we first compute the model's predicted class distribution given a weakly augmented version of a given unlabeled EEG signal:

$$q_i^t = p_c(\widetilde{y}_i^t | X_i^{*t}), \tag{5.9}$$

where $X_i^{*t}$ stands for the weakly augmented version of input $X_i^t$. Then, we use $argmax(q_i^t)$ where $q_i^{*t}$ is no lower than a threshold of 0.95 as a pseudo-label $\widetilde{y}_i'^{*t}$ and apply the cross-entropy loss as below:

$$\mathcal{L}_c^t = \frac{1}{O} \sum_{i=1}^{O} \mathcal{H}(\widetilde{y}_i'^{*t}, p_c(\widetilde{y}_i^t | X_i^t)), \tag{5.10}$$

where $O$ is the number of unlabeled samples in a batch, and $p_c(\widetilde{y}_i^t|X_i^t)$ is the predicted class distribution produced by the epoch view $E_i^t$, which is deemed as the strongly augmented version of input $X_i^t$.

The final loss function of this module is the addition of the above two losses:

$$\mathcal{L}_c = \mathcal{L}_c^s + \mathcal{L}_c^t. \tag{5.11}$$

### 5.3.5 Final Temporal Context based Sleep Stage Classification

The sleep stage classification takes the CAP and epoch views to classify the EEG splits into one of the five stages.

As with the CAP classifier, an LSTM with 128 cells captures the transition pattern between epochs. The design of the stage classifier is as shown in Fig 5.8.

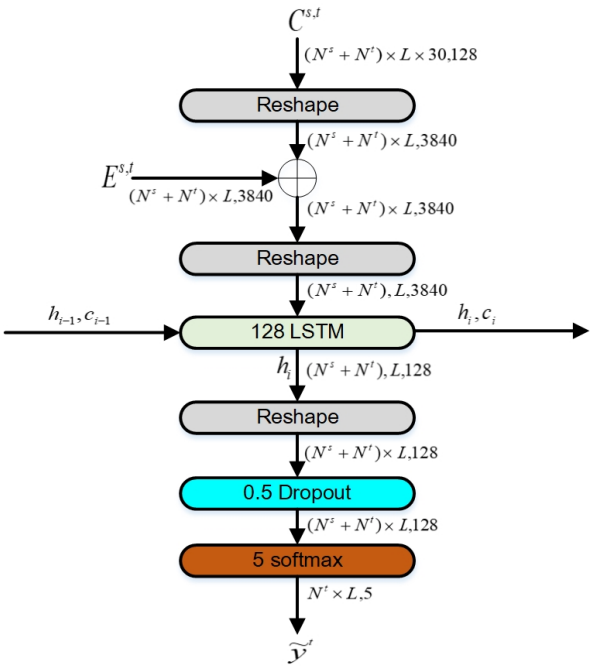


FIGURE 5.8: Design of the stage classifier. The LSTM classifier takes both the epoch-level view and CAP view from each domain as input for classification.

The module takes the addition of the epoch view $E_i$ and the CAP view $C_i$ from both the source and target domains as input and feeds them into an LSTM of 128 units followed by a dropout layer with a keep rate of 0.5. The output is fed into a softmax layer to output the predicted probabilities.

As we only classify the stages in the target dataset and all of them have labels, we apply the similar class-weighted cross-entropy loss function as the source domain CAP classification:

$$\mathcal{L}_e^t = \frac{1}{P}\sum_{i=1}^{P}\mathcal{H}(p_i, p(\widetilde{y}_i^t|X_i^t)), \tag{5.12}$$

$P$ is the number of the target dataset samples in a batch, and $p(\widetilde{y}_i^t|X_i^t)$ is the predicted class distribution produced by the combination of the epoch view and the CAP view from input $X_i^t$ on the target dataset.

## 5.4 Experiments and Results

In this section, we describe two sleep staging datasets, the Montreal Archive of Sleep Studies (MASS) and Sleep-EDF, and introduce the CAP sleep database. Then, we describe the training process and evaluation metrics. After that, the results are compared to the state-of-the-art. To demonstrate the importance of each component in the mode, we provide the ablation study in the end.

### 5.4.1 Datasets

**MASS** [122] contains whole-night recordings of PSG signals from 97 males aged 23 to 63 and 103 females aged from 19 to 57 years. These recordings were collected from 3 different sleep laboratories using eight different research protocols, and the sampling rates are either 256Hz or 512Hz. The recordings are in the European Data Format (EDF)++ and organized into five subsets (SS1-SS5) according to the research protocols. SS1 and SS3 are annotated according to AASM, while the others follow the R&K rule. In this work, we chose SS3 from

cohort 1, which is frequently used in sleep staging research. It contains 28 male and 34 female healthy subjects aged from 20 to 69 years old. The epochs annotated as UNKNOWN are removed, and the rest are labeled into five sleep stages: W, N1, N2, N3, and REM. We use the F4 channel in this work. It is a reformatted signal that subtracts the EEG and EOG channels.

**Sleep-EDF** [120] is another popular dataset for sleep staging. It consists of horizontal EOG, Fpz-Cz, and Pz-Oz EEG at a sampling rate of 100Hz and was collected from 20 healthy Caucasian males and females aged 21 to 35. To conduct fair comparisons with the literature, the Fpz-Cz channel is selected. Each epoch was scored into eight classes according to the R&K rule: W, S1, S2, S3, S4, REM, MOVEMENT, and UNKNOWN. Stages S3 and S4 were merged as N3 to be consistent with MASS, and the epochs annotated as MOVEMENT or UNKNOWN were discarded. In addition, continuous wake epochs longer than 30 minutes outside the sleep period were also ignored in our experiments.

**CAP Sleep Database** [127] is a collection of PSG recordings from 108 subjects where the 16 healthy subjects did not present any neurological disorders and were free of drugs affecting the central nervous system. The rest of the 92 recordings include 40 recordings of patients diagnosed with nocturnal frontal lobe epilepsy, 22 affected by REM behavior disorder, 10 with periodic leg movements, nine insomniacs, five narcoleptics, four affected by sleep-disordered breathing, and two by bruxism. Each record includes EEG signals from at least three channels ((F3 or F4, C3 or C4, and O1 or O2) together with EOG, chin and tibial EMG, airflow, respiratory effort, SaO2, and ECG signals. This database contains sleep stage labels for all epochs and valuable CAP annotations. As with the Sleep-EDF dataset, the sleep stage annotations follow R&K and are processed similarly to the Sleep-EDF dataset.

On the other hand, three sub-phases from A-phase (A1, A2, and A3) are available as CAP annotations. According to the corresponding sleep stage, we created three artificial labels for those EEG signals where CAP annotations are unavailable. They are Unknown Wake (UW) from the Wake stages, Unknown REM (UR) from the REM stages, and Unknown (U) for the remaining unlabeled signals. The sampling rates for recordings vary from 128 to 512 Hz. In this work, we resample all EEG signals from all the datasets into 512 Hz to make them

consistent. The O2 channel is selected in this work as it is the only channel where the CAP labels are available for all 108 subjects.

Table 5.1 summarizes the number of epochs used in the work for each sleep stage from these three datasets.

TABLE 5.1: Number of epochs for each sleep stage from two datasets

| Datasets | Channels | W | N1 | N2 | N3 | REM | Total |
|---|---|---|---|---|---|---|---|
| MASS | F4 | 6231 | 4814 | 29777 | 7653 | 10581 | 59056 |
| Sleep-EDF | Fpz-Cz | 8285 | 2804 | 17799 | 5703 | 7717 | 42308 |
| CAP | O2 | 15919 | 4642 | 39190 | 25732 | 18868 | 104351 |

TABLE 5.2: Comparison between our method and the state-of-the-art on Sleep-EDF Fpz-Cz channel

| **Methods** | **Overall Metrics** | | | **Per-class F1-Score (F1)** | | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | MF1 | $k$ | W | N1 | N2 | N3 | REM |
| IITNet [199] | 84.0 | 77.7 | 0.78 | 87.9 | 44.7 | 88.0 | 85.7 | 82.1 |
| SeqSleepNet+ (FT) [217] | 85.2 | 79.6 | 0.79 | - | - | - | - | - |
| SleepEEGNet [200] | 84.3 | 79.7 | 0.79 | 89.2 | 52.2 | 86.8 | 85.1 | 85.0 |
| DeepSleepNet [65] | 82.0 | 76.9 | 0.76 | 84.7 | 46.6 | 85.9 | 84.8 | 82.4 |
| XSleepNet [204] | 83.9 | 78.7 | 0.77 | - | - | - | - | - |
| TinySleepNet [202] (SOTA) | 85.4 | 80.5 | 0.80 | 90.1 | 51.4 | **88.5** | **88.3** | 84.3 |
| CAPSleepNet (Ours) | **86.2** | **81.0** | **0.81** | **91.0** | **52.5** | 88.4 | 85.7 | **87.3** |

TABLE 5.3: Comparison between our method and the state-of-the-art on MASS F4 channel

| **Methods** | **Overall Metrics** | | | **Per-class F1-Score (F1)** | | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | MF1 | $k$ | W | N1 | N2 | N3 | REM |
| IITNet [199] | 86.6 | 80.8 | 0.80 | 86.1 | 54.4 | 91.3 | 86.0 | 86.2 |
| DeepSleepNet [65] | 86.2 | 81.7 | 0.80 | 87.3 | 59.8 | 90.3 | 81.5 | **89.3** |
| XSleepNet [204] | 85.2 | 80.6 | 0.79 | - | - | - | - | - |
| TinySleepNet [202] (SOTA) | 87.5 | 83.2 | 0.82 | 87.3 | **62.7** | 91.8 | 85.5 | 88.6 |
| CAPSleepNet (Ours) | **88.1** | **83.6** | **0.83** | **87.3** | 62.1 | **92.6** | **88.4** | 87.7 |

### 5.4.2 Training Settings

The $k$-fold cross-validation (CV) is applied to evaluate our model's performance. The $k$ for Sleep-EDF and MASS-SS3 are 20 and 31, respectively, where at least one subject is left out for testing. Furthermore, we split 10% from the training set into a validation set to apply early

stopping. The best checkpoint is saved during training, and the training stops either when it reaches 200 epochs or there is no improvement after 50 epochs. Since we sample from both the source domain and the target domain and feed them into the model simultaneously in mini-batches, the CAP dataset will be split into the same folds depending on which target domain dataset is used.

We use the mini-batch size of 20 and the sequence length of 15. L2 weight decay of $10^{-3}$ is applied to all convolutional layers to prevent overfitting. Adam optimizer at a learning rate of $10^{-4}$ is adopted to supervise the training process. The optimizer's $\beta_1$ and $\beta_2$ are set to 0.9 and 0.999, respectively. The threshold of the gradient clipping is set to 1 to avoid gradient explosion during LSTM training. We randomly shift each epoch along the time axis by a specific range between -3 and 3 seconds during training as data augmentation. On the other hand, we randomly skip the first 0 to 5 epochs to generate new sequences of EEG signals to capture the misalignment between signals and annotations.

Our proposed method was implemented with TensorFlow 1.12. The training was conducted on an NVIDIA GeForce GTX 1080Ti GPU. Our proposed model was trained end-to-end, and each epoch takes around 2 minutes.

### 5.4.3 Evaluation Metrics

Six evaluation metrics are used in our experiments: per-class recall (RE), per-class precision (PR), per-class F1-score (F1), overall accuracy (ACC), macro-averaging F1-score (MF1), and Cohen's Kappa coefficient ($\kappa$).

ACC and MF1 are defined as follows:

$$ACC = \frac{\sum_{i=1}^{K} TP_i}{N_s}, \tag{5.13}$$

$$MF1 = \frac{\sum_{i=1}^{K} F1_i}{N_t}, \tag{5.14}$$

where $TP_i$ and $F1_i$ are the true positives and F1-score of the $i$th stage, respectively, $N_t$ is the total number of sleep stages, which is 5 in our study, and $N_s$ is the total number of test epochs.

Cohen's Kappa Coefficient [152] is to characterize the inter-rater agreement (IRA) level and can be calculated as follows:

$$\kappa = \frac{N_c - N_e}{N_t - N_e}, \tag{5.15}$$

where $N_c$ denotes the number of correctly scored stages, and $N_e$ indicates the expected number of agreements for each stage.

### 5.4.4 Comparison with the State-of-the-art

Table 5.2 and Table 5.3 compare our method with the state-of-the-art across ACC, MF1, $k$, and F1 on the Sleep-EDF Fpz-Cz channel and the MASS F4 channel respectively. We only include the recent techniques that utilize the deep learning model to extract features from the raw single-channel EEGs and evaluate them using independent training and test sets.

Our method scores sleep stages with the ACC and the MF1 of 86.2% and 81.0% on Sleep-EDF, respectively, while the same metrics are 88.1% and 83.6% on MASS. It outperforms the state-of-the-art on both datasets. Our model performs exceptionally well on NREM stages compared to the other approaches due to the leverage of the sleep microstructure learned from CAP information. Also, Cohen's Kappa shows that the agreement between our model and the sleep experts is almost perfect (0.81-1).

We also show the confusion matrix of the Fpz-Cz channel from Sleep-EDF in Table 5.4. From these confusion matrices, the candidate performs worst in the N1 stage and best in the W stage. N1, N3, and REM epochs will likely be misclassified as N2. It may be due to N2 having the most samples in the dataset. The misclassified N2 samples are distributed evenly across other classes, implying that our model discriminates similarly between N2 and other classes. W is most likely misclassified as N1, indicating difficulty identifying sleep onset. The model also effectively distinguishes W from N3 and REM, suggesting that our model

captures temporal context well, as W is not adjacent to N3 and REM in stage transition. The reality that N3 is rarely identified as REM could benefit from the leverage of CAP events.

TABLE 5.4: Confusion matrix from 20 fold cross-validation of Sleep-EDF Fpz-Cz channel and per-class metrics

| **Stages** | **Predicted** | | | | | **Candidate** | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **9355** | 589 | 137 | 15 | 101 | **91.7** | **90.3** | **91.0** |
| N1 | 397 | **1443** | 605 | 10 | 349 | 51.5 | 53.5 | 52.5 |
| N2 | 400 | 425 | **15873** | 579 | 522 | 89.2 | 87.6 | 88.4 |
| N3 | 38 | 6 | 930 | **4729** | 0 | 82.9 | 88.6 | 85.7 |
| REM | 171 | 235 | 582 | 3 | **6726** | 87.2 | 87.4 | 87.3 |

TABLE 5.5: Confusion matrix from 31 fold cross-validation of MASS F4 channel and per-class metrics

| **Stages** | **Predicted** | | | | | **Candidate** | | |
|---|---|---|---|---|---|---|---|---|
| | W | N1 | N2 | N3 | REM | RE | PR | F1 |
| W | **5238** | 644 | 108 | 8 | 188 | 84.7 | 90.1 | 87.3 |
| N1 | 319 | **2977** | 708 | 2 | 790 | 62.1 | 62.0 | 62.1 |
| N2 | 96 | 702 | **27596** | 672 | 668 | **92.8** | **92.4** | **92.6** |
| N3 | 4 | 2 | 1048 | **6599** | 0 | 86.2 | 90.6 | 88.4 |
| REM | 158 | 474 | 397 | 0 | **9508** | 90.2 | 85.2 | 87.7 |

Unlike Sleep-EDF, the model performs best in N2 stages but still worst in N1 according to the confusion matrix of the MASS F4 channel in Table 5.5. However, the prediction on all stages except W is generally better compared to Sleep-EDF, which may be because MASS has a higher sampling rate. Interestingly, the model is likely to be confused by REM and N1, which are not normally close to each other, while it performs well in discriminating N1 and N3. Given that REM and N3 are never misclassified into each other, it might be because subjects in MASS have more dramatic transitions between NREM and REM sleep than those from Sleep-EDF.

We visualized annotated stages along with predicted stages of Fpz-Cz channels from Sleep-EDF dataset in Fig. 5.9 using the first subject's recording. We can see some false positives in the sleep onset phase. However, the drastic changes in sleep stages around the 600th epoch are well captured in our model, demonstrating our model's effectiveness.

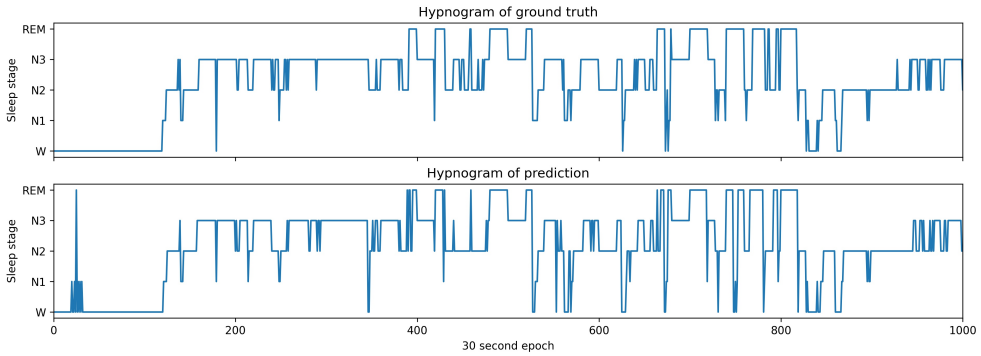


FIGURE 5.9: Hypnograms of subject 1 from Sleep-EDF dataset Fpz-Cz channel. The top is the ground truth, and the bottom is the prediction from our model.

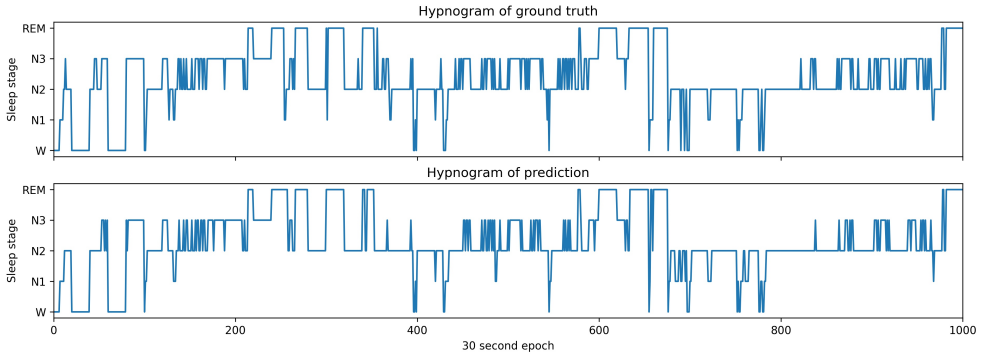


FIGURE 5.10: Hypnograms of subject 1 from MASS dataset F4 channel. The top is the ground truth, and the bottom is the prediction from our model.

A similar visualization from the MASS dataset is provided in Fig. 5.10. Our model performs better than the Sleep-EDF in the sleep onset phase, which could be due to the higher resolution of the MASS dataset. The overall temporal context is well captured, especially the transition between REM and NREM phases, which implies the CAP plays a pivotal role in prediction.

### 5.4.5 Ablation Study

To prove the effective design of the model, we conducted ablation studies to prove the validity of three modules of the method, as shown in Table 5.6. First, we remove the epoch view extractor (EVE) and apply depth-wise convolution directly on the preliminary feature vectors using themselves as convolutional kernels to extract domain styles. Second, we remove CAP

view generation (CVG) using the domain styles as the CAP views. Last, we remove the missing CAP view reconstruction (CVR) by removing the CAP classifier.

TABLE 5.6: Ablation analysis of the performance of the proposed method

| Modules | | | Datasets | Channels | Overall Metrics | | | Per-class F1-Score (F1) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **EVE** | **CVG** | **CVR** | | | **ACC** | **MF1** | $k$ | **W** | **N1** | **N2** | **N3** | **REM** |
| | ✓ | ✓ | Sleep-EDF | Fpz-Cz | 85.5 | 80.0 | 0.80 | 90.7 | 50.6 | 87.9 | 84.9 | 86.1 |
| ✓ | | ✓ | Sleep-EDF | Fpz-Cz | 85.5 | 80.2 | 0.79 | 89.9 | 51.9 | 87.7 | 84.5 | 86.9 |
| ✓ | ✓ | | Sleep-EDF | Fpz-Cz | 84.6 | 79.2 | 0.79 | 89.6 | 50.9 | 87.1 | 81.9 | 86.2 |
| ✓ | ✓ | ✓ | Sleep-EDF | Fpz-Cz | **86.2** | **81.0** | **0.81** | **91.0** | **52.5** | **88.4** | **85.7** | **87.3** |
| | ✓ | ✓ | MASS-SS3 | F4 | 86.6 | 81.8 | 0.80 | 86.1 | 58.9 | 91.3 | 86.2 | 86.3 |
| ✓ | | ✓ | MASS-SS3 | F4 | 86.9 | 82.1 | 0.81 | 85.8 | 60.1 | 91.7 | 86.2 | 86.6 |
| ✓ | ✓ | | MASS-SS3 | F4 | 86.9 | 81.9 | 0.81 | 85.8 | 59.3 | 91.7 | 85.6 | 87.2 |
| ✓ | ✓ | ✓ | MASS-SS3 | F4 | **88.1** | **83.6** | **0.83** | **87.3** | **62.1** | **92.6** | **88.4** | **87.7** |

As per the results, while performance drop can be observed across both Sleep-EDF and MASS datasets, removing the epoch view extraction on Sleep-EDF has the least impact compared to the other modules. Per-class F1 drops most on N1 stages, which indicates that sleep macrostructure may contain discriminative features for N1 stages. When CAP view generation is removed, N3 is impacted most, which implies that adapting sleep microstructure features to sleep microstructure features is essential to N3's classification. When the CAP view reconstruction is removed, the model's performance deteriorates on N3 dramatically due to the lack of supervision of the training process using CAP pseudo labels.

Looking at the MASS dataset, the epoch view also plays a pivotal role in distinguishing N1, aligning with the Sleep-EDF observation. The removal of CAP view generation follows the same pattern on Sleep-EDF, which impacts N3 the most. N1 and N3 drop dramatically when no CAP pseudo labels are involved in the training process.

To summarize, the CAP view improves the model's performance. To effectively generate the CAP views, the epoch view extractor is needed to project the preliminary features into the epoch views to improve the N1 stages' classification. Adapting epoch views into the CAP views by fusion the epoch views with the domain styles can better transfer sleep macrostructure features to sleep microstructure features. The CAP classifier is indispensable in using the CAP pseudo labels in representation learning.

### 5.4.6 Computational Cost Analysis

While detailed computational cost analysis is often crucial for deployment considerations, our focus in this work is on demonstrating the effectiveness of incorporating sleep microstructure (CAP) features for improved sleep staging accuracy. The computational overhead introduced by our multi-view architecture is justified by the significant performance improvements achieved.

Our model's training time of approximately 2 minutes per epoch on an NVIDIA GeForce GTX 1080Ti GPU is comparable to other deep learning approaches in sleep staging literature. The additional computational cost primarily stems from:

- Domain style extraction module requiring dynamic convolution operations
- CAP view generation and pseudo-labeling process
- Dual LSTM classifiers for both CAP and sleep stage prediction

However, these computational costs are offset by the clinical value of improved sleep staging accuracy, particularly for distinguishing challenging sleep stages like N1 and N3. The model's inference time remains practical for clinical applications, and the improved diagnostic accuracy justifies the additional computational requirements. Future work could explore model compression techniques or efficient architectures to reduce computational overhead while maintaining the performance gains from sleep microstructure integration.

## 5.5 Conclusion

This chapter presents an end-to-end domain adaptation deep learning method for sleep microstructure-based sleep staging. The model extracts the domain styles that contain multiple factors from both the source and the target datasets, then the domain styles adapt the epoch-level view to the CAP views. A CAP classifier transfers the CAP features from the source dataset to the target dataset on a different EEG channel. A weakly augmented version of signals from the target dataset is used to reconstruct the missing CAP view in the target dataset

during training. The epoch view and CAP view are combined to predict the stages in the end. Experimental results on two widely used public datasets demonstrate that our proposed method outperforms the state-of-the-art deep learning-based approaches by a significant margin, proving that collecting CAP labels for sleep studies is essential.

CHAPTER 6

# Conclusions and Future Work

## 6.1 Summary

The landscape of sleep medicine has been fundamentally transformed by advances in computational analysis of physiological data. Modern sleep monitoring technologies generate unprecedented volumes of neurophysiological recordings, characterized by improved temporal resolution, larger patient cohorts, and enhanced spatial precision. However, this wealth of data has also revealed substantial inter-patient variability that demands sophisticated computational approaches to extract meaningful clinical insights.

This dissertation commenced with a thorough literature review in Chapter 2, synthesizing current developments across signal acquisition, analytical methodologies, and clinical applications in sleep research. The review traced the field's evolution from labor-intensive manual scoring to automated staging systems, examined progress in sleep disorder detection algorithms, and evaluated emerging alternative monitoring approaches. Additionally, it highlighted the growing importance of publicly available datasets in enabling reproducible research. Through this analysis, several key trends emerged that are reshaping sleep signal analytics: the convergence of multiple physiological data streams, the development of less intrusive monitoring solutions, the widespread adoption of machine learning techniques, and an increasing emphasis on personalized treatment approaches. The review concluded by identifying critical research gaps, particularly the need for hybrid methodologies that combine traditional signal processing expertise with deep learning capabilities, investigation of novel non-invasive sensors, development of algorithms robust to individual differences and

environmental variations, and deeper exploration of connections between sleep patterns and broader health outcomes.

The subsequent chapters tackled three fundamental challenges that have emerged as computational methods become increasingly central to sleep medicine practice. In Chapter 3, we examined the transition from conventional signal processing to neural network-based approaches in sleep analysis. The chapter introduced a novel deep learning architecture that incorporates hierarchical convolutional components to emulate frequency-domain transformations within a transformer framework. This approach represents an attempt to bridge traditional signal theory with automated feature learning, resulting in more efficient training procedures for neural networks. The methodology proved capable of handling large patient datasets across multiple EEG channels, laying important groundwork for generalizable sleep classification systems. Our experiments demonstrated that integrating established signal processing principles with modern transformer architectures significantly enhances feature extraction from single-channel EEG data, though certain limitations in cross-dataset generalization became apparent during testing.

Chapter 4 shifted focus toward the extraction and clinical application of sleep-related biomarkers from large-scale clinical databases, with particular attention to predictive modeling of insomnia-related EEG patterns. Working with an extensive patient dataset enabled robust model development for sleep stage classification while revealing novel clinical associations. The validation studies with insomnia patients confirmed that extracted neurophysiological features possess genuine predictive value, suggesting potential applications in precision medicine approaches. However, the analytical framework outlined in this chapter, while promising for developing patient-specific predictive tools, requires extensive validation across diverse patient populations and medical conditions before clinical implementation becomes feasible.

The final research component in Chapter 5 evaluated advanced computational strategies for sleep stage prediction through detailed neurophysiological pattern analysis. This work introduced novel approaches to improve classification accuracy via comprehensive examination of signal microarchitecture. Through systematic evaluation of existing methodologies, we established performance benchmarks for future sleep staging research while identifying areas

where detailed signal analysis shows particular promise for disease identification and patient stratification. The results suggest that incorporating fine-grained neurophysiological analysis could significantly advance both diagnostic capabilities and our understanding of sleep-related pathophysiology.

## 6.2 Recent Advances in Sleep Signal Analytics

The field has experienced rapid evolution since the completion of the core research presented in this thesis. Developments in deep learning architectures, multimodal data integration, clinical translation strategies, and sensor technologies have not only addressed some limitations identified in earlier work but have also opened entirely new research directions that contextualize this thesis within the broader trajectory of sleep research.

Deep learning approaches have become increasingly sophisticated since 2023, with architectures specifically designed for sleep-related applications showing marked improvements over earlier general-purpose models. Siamese-based networks [218], hybrid CNN-transformer architectures [219], and convolutional recurrent transformers [220] have demonstrated superior performance compared to the CNN and RNN approaches that dominated earlier research. These developments represent a natural evolution of the computational methods explored in this thesis, though they also highlight the rapid pace of change in this field.

The integration of multiple data modalities has emerged as particularly promising, with researchers demonstrating that combining EEG with peripheral physiological signals can substantially improve sleep staging accuracy. Notable contributions include infrared-based monitoring systems that fuse visual and bioelectrical data [221], lightweight architectures optimized for multi-channel polysomnography [222], and multimodal fusion networks that leverage hybrid embedding approaches [223]. These studies collectively emphasize the value of cross-modal feature learning, extending beyond the single-channel EEG focus of much earlier work and providing a foundation for more comprehensive sleep assessment methodologies.

Clinical translation has received increased attention, with researchers addressing the interpretability challenges that have historically limited deep learning adoption in clinical settings. Important progress includes interpretable classification schemes based on layer-wise relevance propagation [224] and multiclass activation mapping approaches [225]. Furthermore, investigations into channel contribution patterns in deep learning-based scoring systems have provided valuable insights into optimal multi-channel model design [226]. These advances directly address the clinical applicability concerns identified throughout this thesis, though significant validation work remains before widespread clinical deployment becomes feasible.

Novel sensing technologies have also made substantial progress, offering alternatives to traditional PSG that could dramatically improve patient comfort and accessibility. Ultra-wideband radar systems integrated into consumer devices [227], contactless camera-based staging systems [228], and advanced wearable technologies with improved cross-population generalizability [229] represent significant innovations in patient-centered monitoring. While these approaches show promise, they also introduce new challenges around validation, standardization, and integration with existing clinical workflows.

These recent developments collectively push sleep signal analytics toward more robust, interpretable, and clinically relevant solutions. While this thesis contributes to foundational aspects of sleep staging and insomnia detection, the emerging research directions highlight numerous opportunities for extending and enhancing the proposed methodologies. The continued vitality and rapid evolution of the field suggest that the frameworks developed here provide a solid foundation for future advances.

## 6.3 Future Work

Despite addressing several fundamental challenges in advanced sleep signal analysis, significant opportunities for methodological improvement remain. Modern neural network approaches, while showing considerable promise, face practical barriers to clinical implementation due to substantial data requirements and interpretation challenges. The domain adaptation frameworks explored in Chapter 4, combined with supervised learning techniques,

offer potential solutions, though additional research is needed as computational resources continue to expand. Our findings suggest that current data limitations may constrain methodological effectiveness, a hypothesis that warrants further investigation as technology advances. Moreover, developing transparent machine learning approaches remains essential for meaningful clinical integration.

Accessibility and standardization of analytical methods represents a critical challenge for advancing sleep medicine research. Chapter 5 highlighted the substantial technical infrastructure required for methodology validation, particularly for biosignal pattern recognition and the comparative feature studies discussed in Chapter 4. The field would benefit significantly from unified coding frameworks and repositories supporting multiple programming environments, including both R and Python implementations. Such standardization would facilitate cross-dataset methodology application and enable broader implementation across research groups. While some sleep research domains currently lack standardized tools, developments in biomedical signal processing standardization provide valuable models for future efforts.

The methodological framework presented addresses multiple aspects of physiological data analysis workflows, emphasizing comprehensive feature extraction, patient stratification, comparative model evaluation, and signal preprocessing optimization for clinical prediction accuracy. The approaches detailed across Chapters 3, 4, and 5 demonstrate adaptability to emerging technologies, providing flexible frameworks for future investigations. However, their current application primarily to single-channel neurophysiological recordings across limited datasets represents a significant constraint. Future research incorporating multiple physiological measurements could provide deeper insights into disease mechanisms and their complex interactions. This integrated analysis approach might reveal novel biomarker combinations, potentially improving disease outcome prediction precision.

The research process revealed persistent challenges in extracting meaningful information from large sleep medicine datasets. While signal complexity and granularity have improved substantially, patient cohort sizes often remain insufficient for comprehensive disease analysis, particularly for rare conditions or specific patient subgroups. However, our work in Chapter 4 demonstrated that large patient populations can be effectively leveraged through computational

adaptation strategies for data enrichment. Developing pre-trained analytical models that link physiological features to clinical outcomes represents a promising direction, especially given anticipated technological advances. Despite current limitations, this research underscores the significant potential of these approaches to enhance physiological understanding and guide personalized therapeutic interventions.

The methodological innovations presented extend well beyond sleep medicine applications. The exponential growth in data generation across multiple disciplines creates opportunities for both personalization and large-scale model development. Applications span consumer analytics, multimedia recommendation systems, and behavioral prediction models, among others. These diverse implementations require domain expertise for effective feature extraction (as demonstrated in Chapter 3) and utilize population heterogeneity in large datasets for appropriate stratification (employing methods similar to those developed in Chapter 4). The emergence of advanced language processing systems exemplifies the potential of expanded data availability and computational capacity, enabling sophisticated interactive systems and multimedia analysis capabilities. Our evaluation approach in Chapter 5 demonstrates the importance of understanding model performance characteristics and application-specific requirements.

The advancement of sleep medicine continues to depend heavily on sophisticated computational capabilities and ongoing technological innovation. Current progress in understanding and treating sleep disorders relies increasingly on integrating multiple physiological data streams with advanced statistical and computational methods. This dissertation contributes to sleep medicine advancement while establishing frameworks for future developments in biomedical data science. The interdisciplinary approaches and methodologies developed across all chapters provide valuable contributions to data-driven applications in scientific domains, particularly regarding emerging technologies and expanding computational capabilities. The frameworks and methodologies presented establish a foundation for continued advancement in both sleep medicine research and broader applications of computational analysis in biomedical investigation. As techniques continue to advance and our understanding deepens, future work should focus on addressing these challenges while maintaining the rigorous standards

necessary for clinical applications. Particular attention should be directed toward developing more robust, accessible, and user-friendly systems that can effectively bridge the gap between laboratory research and practical clinical implementation.

# Bibliography


[1] Katharina Wulff et al. 'Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease'. In: *Nature Reviews Neuroscience* 11.8 (2010), p. 589.

[2] Le Shi et al. 'Sleep disturbances increase the risk of dementia: A systematic review and meta-analysis'. In: *Sleep Medicine Reviews* 40 (2018), pp. 4–16. ISSN: 1087-0792. DOI: https://doi.org/10.1016/j.smrv.2017.06.010. URL: https://www.sciencedirect.com/science/article/pii/S1087079217300114.

[3] Wei Lin et al. 'Risk of neurodegenerative diseases in patients with sleep disorders: A nationwide population-based case-control study'. In: *Sleep Medicine* 107 (2023), pp. 289–299. ISSN: 1389-9457. DOI: https://doi.org/10.1016/j.sleep.2023.05.014. URL: https://www.sciencedirect.com/science/article/pii/S1389945723001831.

[4] Stephanie D Davis, Ernst Eber, Anastassios C Koumbourlis et al. *Diagnostic Tests in Pediatric Pulmonology*. Springer, 2015.

[5] M. G. Terzano et al. 'The Cyclic Alternating Pattern As a Physiologic Component of Normal NREM Sleep'. In: *Sleep* 8.2 (1985), pp. 137–145. ISSN: 01618105. DOI: 10.1093/sleep/8.2.137.

[6] Z Liang and T Nishimura. 'Are wearable EEG devices more accurate than fitness wristbands for home sleep Tracking? Comparison of consumer sleep trackers with clinical devices'. In: *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. 2017, pp. 1–5. DOI: 10.1109/GCCE.2017.8229188.

[7] Z Liang and M A C Martell. 'Considering interpersonal differences in validating wearable sleep-tracking technologies'. In: *2017 Tenth International Conference on*

*Mobile Computing and Ubiquitous Network (ICMU)*. Oct. 2017, pp. 1–4. DOI: 10.23919/ICMU.2017.8330096.

[8] Evan D Chinoy et al. 'Performance of seven consumer sleep-tracking devices compared with polysomnography'. In: *Sleep* 44.5 (May 2021), zsaa291. ISSN: 0161-8105. DOI: 10.1093/sleep/zsaa291. URL: https://doi.org/10.1093/sleep/zsaa291.

[9] Tauhidur Rahman et al. 'Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar'. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. 2015, pp. 39–50.

[10] Hong Hong et al. 'Noncontact Sleep Stage Estimation Using a CW Doppler Radar'. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* (2018).

[11] J Liu et al. 'Monitoring Vital Signs and Postures During Sleep Using WiFi Signals'. In: *IEEE Internet of Things Journal* (2018), p. 1. DOI: 10.1109/JIOT.2018.2822818.

[12] Sirvan Khalighi et al. 'Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels'. In: *Expert Systems with Applications* 40.17 (2013), pp. 7046–7059.

[13] Anna Krakovská and Krist´ Mezeiová. 'Automatic sleep scoring: A search for an optimal combination of measures'. In: *Artificial intelligence in medicine* 53.1 (2011), pp. 25–33.

[14] Jun Shi et al. 'Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning'. In: *Journal of neuroscience methods* 254 (2015), pp. 94–101.

[15] V C Figueroa Helland et al. 'Investigation of an automatic sleep stage classification by means of multiscorer hypnogram'. In: *Methods of Information in Medicine* 49.05 (2010), pp. 467–472.

[16] Reza Boostani, Foroozan Karimzadeh and Mohammad Nami. *A comparative review on sleep stage classification methods in patients and healthy individuals*. 2017. DOI: 10.1016/j.cmpb.2016.12.004.

[17] Arthur S Walters and David B Rye. ‘Review of the relationship of restless legs syndrome and periodic limb movements in sleep to hypertension, heart disease, and stroke’. In: *Sleep* 32.5 (2009), pp. 589–597.

[18] Tetyana Kendzerska et al. ‘Incident cardiovascular events and death in individuals with restless legs syndrome or periodic limb movements in sleep: a systematic review’. In: *Sleep* 40.3 (2017).

[19] F Mendonca et al. ‘A Review of Obstructive Sleep Apnea Detection Approaches’. In: *IEEE Journal of Biomedical and Health Informatics* (2018), p. 1. ISSN: 2168-2194. DOI: 10.1109/JBHI.2018.2823265.

[20] Mehrnaz Shokrollahi and Sridhar Krishnan. ‘A review of sleep disorder diagnosis by electromyogram signal analysis’. In: *Critical Reviews™ in Biomedical Engineering* 43.1 (2015).

[21] Ernst Niedermeyer and F H Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[22] Alireza Talesh et al. ‘Balance-energy of resting state network in obsessive-compulsive disorder’. In: *Scientific Reports* 13 (2023). DOI: 10.1038/s41598-023-37304-9.

[23] E Malaekah. ‘Automated sleep stage detection and classification of sleep disorders’. PhD thesis. RMIT University, 2016.

[24] Fabrizio De Carli et al. ‘Quantitative analysis of sleep EEG microstructure in the time–frequency domain’. In: *Brain Research Bulletin* 63.5 (2004), pp. 399–405. ISSN: 0361-9230. DOI: https://doi.org/10.1016/j.brainresbull.2003.12.013. URL: https://www.sciencedirect.com/science/article/pii/S0361923004000929.

[25] Ali Torabi and mohammad reza Daliri. ‘Applying nonlinear measures to the brain rhythms: an effective method for epilepsy diagnosis’. In: *BMC Medical Informatics and Decision Making* 21 (2021). DOI: 10.1186/s12911-021-01631-6.

[26] Robert Oostenveld and Peter Praamstra. ‘The five percent electrode system for high-resolution EEG and ERP measurements’. In: *Clinical neurophysiology* 112.4 (2001), pp. 713–719.

[27] Charles Pollak, Michael J Thorpy and Jan Yager. *The encyclopedia of sleep and sleep disorders*. InfoBase publishing, 2010.

[28] Richard Berry. *Fundamentals of Sleep Medicine*. 1st ed. Elsevier Health Sciences, 2012. ISBN: 9781437703269.

[29] D G Robertson et al. 'Research methods in biomechanics . Champaign, IL: Human Kinetics'. In: *Rodrigues JM, Lu's AL, Lobato JV, Pinto MV, Lopes MA, Freitas M, Geuna S, Santos JD, Maur'cio AC,(2005a)." Determination of the intracellular Ca2+ concentration in the N1E-115 neuronal cell line in perspective of its use for peripheric nerve reg* 15 (2004), pp. 455–465.

[30] M D ECG-simplified Aswini Kumar. *LifeHugger*. Tech. rep. Retrieved 2010-02-11, 2010.

[31] Tabris Salsa, Francisco Veiga and Maria Pina. 'Oral Controlled-Release Dosage Forms. I. Cellulose Ether Polymers in Hydrophilic Matrices'. In: *Drug Development and Industrial Pharmacy - DRUG DEVELOP IND PHARM* 23 (1997), pp. 929–938. DOI: 10.3109/03639049709148697.

[32] H Webb Agnew, Wilse B Webb and Robert L Williams. 'The First Night Effect: An Eeg Studyof Sleep'. In: *Psychophysiology* 2.3 (1966), pp. 263–266.

[33] Jaspal Singh and R K Sharma. 'Making sleep study instrumentation more unobtrusive'. In: *IEEE Instrumentation & Measurement Magazine* 21.1 (2018), pp. 50–53.

[34] S Eyal and A Baharav. 'Sleep insights from the finger tip: How photoplethysmography can help quantify sleep'. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.274-197.

[35] Y Liu et al. 'The analysis of college students' sleep status'. In: *2017 Chinese Automation Congress (CAC)*. 2017, pp. 4501–4504. DOI: 10.1109/CAC.2017.8243573.

[36] Y Yoshida et al. 'Development of sleep-wake estimation algorithm using the wrist acceleration sensor'. In: *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. 2017, pp. 1–2. DOI: 10.1109/GCCE.2017.8229315.

[37] Veronica Cabanas-Sanchez et al. 'Automated algorithms for detecting sleep period time using a multi-sensor pattern-recognition activity monitor from 24-h free-living

data in older adults'. In: *Physiological Measurement* (2018). URL: http://iopscience.iop.org/10.1088/1361-6579/aabf26.

[38] G S Umemura et al. 'Comparison of sleep parameters assessed by actigraphy of healthy young adults from a small town and a megalopolis in an emerging country'. In: *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*. 2017, pp. 200–203. DOI: 10.1109/HIC.2017.8227619.

[39] S Zhou et al. 'Analysis of Health and Physiological Index Based on Sleep and Walking Steps by Wearable Devices for the Elderly'. In: *2017 IEEE 10th Conference on Service-Oriented Computing and Applications (SOCA)*. 2017, pp. 245–250. DOI: 10.1109/SOCA.2017.42.

[40] S K Rocknathan et al. 'Empirical evaluation of consumer EEG and actigraphy devices for home-based sleep assessment'. In: *2017 International Conference on Orange Technologies (ICOT)*. 2017, pp. 49–52. DOI: 10.1109/ICOT.2017.8336086.

[41] A Khademi et al. 'Toward personalized sleep-wake prediction from actigraphy'. In: *IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. 2018, pp. 414–417. DOI: 10.1109/BHI.2018.8333456.

[42] Tian Hao, Guoliang Xing and Gang Zhou. 'iSleep: unobtrusive sleep quality monitoring using smartphones'. In: *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM. 2013, p. 4.

[43] L Montanini et al. 'Smartphone as unobtrusive sensor for real-time sleep recognition'. In: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. Jan. 2018, pp. 1–4. DOI: 10.1109/ICCE.2018.8326220.

[44] Zhenyu Chen et al. 'Unobtrusive sleep monitoring using smartphones'. In: *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*. IEEE. 2013, pp. 145–152.

[45] S Vhaduri and C Poellabauer. 'Impact of different pre-sleep phone use patterns on sleep quality'. In: *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. Mar. 2018, pp. 94–97. DOI: 10.1109/BSN.2018.8329667.

[46] Cheng Yang et al. ‘Sleep apnea detection via depth video and audio feature learning’. In: *IEEE Transactions on Multimedia* 19.4 (2017), pp. 822–835.

[47] T Nochino, Y Ohno and S Okada. ‘Development of noncontact respiration monitoring method with web-camera during sleep’. In: *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. 2017, pp. 1–2. DOI: 10.1109/GCCE.2017.8229408.

[48] X Yang et al. ‘Monitoring of Patients Suffering from REM Sleep Behavior Disorder’. In: *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* (2018), p. 1. ISSN: 2469-7249. DOI: 10.1109/JERM.2018.2827705.

[49] A Alivar et al. ‘Motion Detection in Bed-Based Ballistocardiogram to Quantify Sleep Quality’. In: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8255014.

[50] Allan Rechtschaffen. ‘A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects’. In: *Public Health Service* (1968).

[51] American Academy of Sleep Medicine et al. ‘The AASM manual for the scoring of sleep and associated events: Rules’. In: *Terminology and Technical Specifications* (2007), pp. 48–49.

[52] William O Tatum. ‘Ellen R. Grass Lecture: Extraordinary EEG’. In: *The Neurodiagnostic Journal* 54.1 (2014), pp. 3–21.

[53] Seral Özsen. ‘Classification of sleep stages using class-dependent sequential feature selection and artificial neural network’. In: *Neural Computing and Applications* 23.5 (2013), pp. 1239–1250.

[54] W W Flemons et al. ‘Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research’. In: *Sleep* 22.5 (1999), pp. 667–689.

[55] Jose M Marin et al. ‘Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study’. In: *The Lancet* 365.9464 (2005), pp. 1046–1053.

[56] Richard B Berry et al. 'The AASM manual for the scoring of sleep and associated events'. In: *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine* (2012).

[57] Carl Stepnowsky et al. 'Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters'. In: *Sleep medicine* 14.11 (2013), pp. 1199–1207.

[58] Shing-Tai Pan et al. 'A transition-constrained discrete hidden Markov model for automatic sleep staging'. In: *Biomedical engineering online* 11.1 (2012), p. 52.

[59] Alexandra Piryatinska et al. 'Optimal channel selection for analysis of EEG-sleep patterns of neonates'. In: *Computer methods and programs in biomedicine* 106.1 (2012), pp. 14–26.

[60] Gi-Ren Liu et al. 'Diffuse to fuse EEG spectra–intrinsic geometry of sleep dynamics for classification'. In: *arXiv preprint arXiv:1803.01710* (2018).

[61] Emina Alickovic and Abdulhamit Subasi. 'Ensemble SVM Method for Automatic Sleep Stage Classification'. In: *Submitted to IEEE TIM* ().

[62] Z Zhang et al. 'Different bands of sleep EEG analysis based on the multiscale Jenson-Shannon Divergence'. In: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 2017, pp. 1–5. DOI: 10.1109/CISP-BMEI.2017.8302218.

[63] Orestis Tsinalis et al. 'Automatic sleep stage scoring with single-channel EEG using convolutional neural networks'. In: *arXiv preprint arXiv:1610.01683* (2016).

[64] Arnaud Sors et al. 'A convolutional neural network for sleep stage scoring from raw single-channel eeg'. In: *Biomedical Signal Processing and Control* 42 (2018), pp. 107–114.

[65] Akara Supratak et al. 'DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.11 (2017), pp. 1998–2008.

[66] Junming Zhang and Yan Wu. 'A New Method for Automatic Sleep Stage Classification'. In: *IEEE transactions on biomedical circuits and systems* 11.5 (2017), pp. 1097–1110.

[67] Kaare Mikkelsen and Maarten de Vos. 'Personalizing deep learning models for automatic sleep staging'. In: *arXiv preprint arXiv:1801.02645* (2018).

[68] S Chambon et al. 'A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.4 (2018), pp. 758–769. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2018.2813138.

[69] Hubert Cecotti and Axel Graeser. 'Convolutional neural network with embedded Fourier transform for EEG classification'. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE. 2008, pp. 1–4.

[70] Yu-Liang Hsu et al. 'Automatic sleep stage recurrent neural classifier using energy features of EEG signals'. In: *Neurocomputing* 104 (2013), pp. 105–114.

[71] Intan Nurma Yulita, Mohamad Ivan Fanany and Aniati Murni Arymurthy. 'Combining deep belief networks and bidirectional long short-term memory: Case study: Sleep stage classification'. In: *Electrical Engineering, Computer Science and Informatics (EECSI), 2017 4th International Conference on*. IEEE. 2017, pp. 1–6.

[72] Z Zhang and C Guan. 'An accurate sleep staging system with novel feature generation and auto-mapping'. In: *2017 International Conference on Orange Technologies (ICOT)*. 2017, pp. 214–217. DOI: 10.1109/ICOT.2017.8336079.

[73] American Academy of Sleep Medicine et al. 'International classification of sleep disorders'. In: *Diagnostic and coding manual* (2005), pp. 51–55.

[74] Angela Louise D'Rozario et al. 'Impaired neurobehavioural performance in untreated obstructive sleep apnea patients using a novel standardised test battery'. In: *Frontiers in Surgery* 5 (2018), p. 35.

[75] Kelly A Loffler et al. 'Effect of obstructive sleep apnea treatment on renal function in patients with cardiovascular disease'. In: *American journal of respiratory and critical care medicine* 196.11 (2017), pp. 1456–1462.

[76] Alia Al-Alawi et al. 'Prevalence, risk factors and impact on daytime sleepiness and hypertension of periodic leg movements with arousals in patients with obstructive sleep apnea'. In: *Journal of Clinical Sleep Medicine* 2.03 (2006), pp. 281–287.

[77] M Deviaene et al. ‘Assessing cardiovascular comorbidities in sleep apnea patients using SpO2’. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.232-224.

[78] G C Gutierrez-Tobal et al. ‘Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity from at-Home Oximetry Recordings’. In: *IEEE Journal of Biomedical and Health Informatics* (2018), p. 1. ISSN: 2168-2194. DOI: 10.1109/JBHI.2018.2823384.

[79] J Lázaro et al. ‘Pulse photoplethysmography derived respiration for obstructive sleep apnea detection’. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.273-229.

[80] Sunil Kumar Telagamsetti and Vivek Kanhangad. ‘Gabor Filter-based 1D-Local Phase Descriptors for Obstructive Sleep Apnea Detection using Single-lead ECG’. In: *IEEE Sensors Letters* (2018).

[81] Y Li et al. ‘Sliding Trend Fuzzy Approximate Entropy as a Novel Descriptor of Heart Rate Variability in Obstructive Sleep Apnea’. In: *IEEE Journal of Biomedical and Health Informatics* (2018), p. 1. ISSN: 2168-2194. DOI: 10.1109/JBHI.2018.2790968.

[82] V Bari et al. ‘Impact of nonstationarities on short heart rate variability recordings during obstructive sleep apnea’. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.203-166.

[83] L Burattini et al. ‘Overnight T-wave alternans in sleep apnea patients’. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.235-086.

[84] Sharadha Kolappan et al. ‘A low-cost approach for wide-spread screening of periodic leg movements related to sleep disorders’. In: *Humanitarian Technology Conference (IHTC), 2017 IEEE Canada International*. IEEE. 2017, pp. 105–108.

[85] Daniel Waltisberg et al. ‘Detecting disordered breathing and limb movement using in-bed force sensors’. In: *IEEE journal of biomedical and health informatics* 21.4 (2017), pp. 930–938.

[86] N Du et al. 'ApneaRadar: A 24GHz Radar-Based Contactless Sleep Apnea Detection System'. In: *2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST)*. 2017, pp. 372–376. DOI: 10.1109/ICFST.2017.8210539.

[87] Heinrich Garn et al. '3D detection of periodic limb movements in sleep'. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 427–430.

[88] S Milici et al. 'Wireless Wearable Magnetometer-Based Sensor for Sleep Quality Monitoring'. In: *IEEE Sensors Journal* 18.5 (2018), pp. 2145–2152. ISSN: 1530-437X. DOI: 10.1109/JSEN.2018.2791400.

[89] S Gopal and T A Devi. 'Obstructive sleep apnea detection from ECG signal using neuro-fuzzy classifier'. In: *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. July 2017, pp. 910–915. DOI: 10.1109/ICICICT1.2017.8342686.

[90] Margot Deviaene et al. 'Automatic Screening of Sleep Apnea Patients Based on the SpO2 Signal'. In: *IEEE Journal of Biomedical and Health Informatics* (2018).

[91] Madeleine M Grigg-Damberger. 'The AASM Scoring Manual four years later'. In: *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine* 8.3 (2012), p. 323.

[92] Maowei Cheng et al. 'Recurrent neural network based classification of ecg signal features for obstruction of sleep apnea detection'. In: *Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on*. Vol. 2. IEEE. 2017, pp. 199–202.

[93] Rim Haidar, Irena Koprinska and Bryn Jeffries. 'Sleep Apnea Event Detection from Nasal Airflow Using Convolutional Neural Networks'. In: *International Conference on Neural Information Processing*. Springer. 2017, pp. 819–827.

[94] S Rezaei et al. 'Diagnosis of sleep apnea by evaluating points distribution in poincare plot of RR intervals'. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.158-398.

[95] S Kye et al. 'Detecting periodic limb movements in sleep using motion sensor embedded wearable band'. In: *2017 IEEE International Conference on Systems, Man,*

*and Cybernetics (SMC)*. Oct. 2017, pp. 1087–1092. DOI: 10.1109/SMC.2017.8122756.

[96] Daniel J Buysse et al. 'Recommendations for a standard research assessment of insomnia'. In: *Sleep* 29.9 (2006), pp. 1155–1173.

[97] Børge Sivertsen et al. 'The epidemiology of insomnia: Associations with physical and mental health.: The HUNT-2 study'. In: *Journal of psychosomatic research* 67.2 (2009), pp. 109–116.

[98] Charles M Morin and Ruth Benca. 'Chronic insomnia'. In: *The Lancet* 379.9821 (2012), pp. 1129–1141.

[99] Saad M Alsaadi et al. 'Detecting insomnia in patients with low back pain: accuracy of four self-report sleep measures'. In: *BMC musculoskeletal disorders* 14.1 (2013), p. 196.

[100] Christopher B Miller et al. 'Clusters of insomnia disorder: an exploratory cluster analysis of objective sleep parameters reveals differences in neurocognitive functioning, quantitative EEG, and heart rate variability'. In: *Sleep* 39.11 (2016), pp. 1993–2004.

[101] Christopher B Miller et al. 'An objective short sleep insomnia disorder subtype is associated with reduced brain metabolite concentrations in vivo: a preliminary magnetic resonance spectroscopy assessment'. In: *Sleep* 40.11 (2017), zsx148.

[102] Orfeu M Buxton and Enrico Marcelli. 'Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States'. In: *Social Science & Medicine* 71.5 (2010), pp. 1027–1036.

[103] Margaret K Vernon et al. 'Measurement of non-restorative sleep in insomnia: A review of the literature'. In: *Sleep Medicine Reviews* 14.3 (2010), pp. 205–212.

[104] Mostafa Shahin et al. 'Deep learning and insomnia: assisting clinicians with their diagnosis'. In: *IEEE Journal of Biomedical and Health Informatics* 21.6 (2017), pp. 1546–1553.

[105] Lamana Mulaffer et al. 'Comparing two insomnia detection models of clinical diagnosis techniques'. In: *IEEE Conference of Engineering in Medicine and Biology Society*. 2017, pp. 3749–3752.

[106] Vernon J Lawhern et al. 'EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces'. In: *Journal of Neural Engineering* 15.5 (July 2018), p. 56013. DOI: 10.1088/1741-2552/aace8c. URL: https://dx.doi.org/10.1088/1741-2552/aace8c.

[107] Sebastian Stober et al. 'Deep feature learning for EEG recordings'. In: *arXiv preprint arXiv:1511.04306* (2015).

[108] Robin Tibor Schirrmeister et al. 'Deep learning with convolutional neural networks for EEG decoding and visualization'. In: *Human brain mapping* 38.11 (2017), pp. 5391–5420.

[109] M A B Altaf and W Saadeh. 'A 0.21 $\mu$J patient-specific REM/Non-REM sleep classifier for Alzheimer patients'. In: *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 2017, pp. 1–4. DOI: 10.1109/BIOCAS.2017.8325225.

[110] S Q Zheng et al. 'Experiments show that a light that can improve sleep quality in terms of hormone concentration'. In: *2017 14th China International Forum on Solid State Lighting: International Forum on Wide Bandgap Semiconductors China (SSLChina: IFWS)*. 2017, pp. 101–104. DOI: 10.1109/IFWS.2017.8245984.

[111] Z Weng et al. 'Psychological and physiological influences of CCTs on young people before sleep'. In: *2017 14th China International Forum on Solid State Lighting: International Forum on Wide Bandgap Semiconductors China (SSLChina: IFWS)*. 2017, pp. 105–108. DOI: 10.1109/IFWS.2017.8245985.

[112] X Peng et al. 'Sleep-deprived fatigue pattern analysis using large-scale selfies from social media'. In: *2017 IEEE International Conference on Big Data (Big Data)*. 2017, pp. 2076–2084. DOI: 10.1109/BigData.2017.8258154.

[113] C Ma, W Xie and J Sun. 'Intelligent Control of Electric Appliances Based on Sleep Perception'. In: *2018 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS)*. Jan. 2018, pp. 564–567. DOI: 10.1109/ICITBS.2018.00148.

[114] C Torres et al. 'A Multiview Multimodal System for Monitoring Patient Sleep'. In: *IEEE Transactions on Multimedia* (2018), p. 1. ISSN: 1520-9210. DOI: 10.1109/TMM.2018.2829162.

[115] O S Eyobu et al. 'A real-time sleeping position recognition system using IMU sensor motion data'. In: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. 2018, pp. 1–2. DOI: 10.1109/ICCE.2018.8326209.

[116] A Opaliński et al. 'The system for integration of heterogeneous data sources in the domain of Obstructive Sleep Apnea'. In: *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*. 2017, pp. 1–6. DOI: 10.1109/BESC.2017.8256370.

[117] M Sitarek et al. 'Central haemodynamic variability during sleep in subjects with and without atrial fibrillation'. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.295-234.

[118] K Q Lepage and S Vijayan. 'System identification for sleep-mediated, stimulation-enhanced memory consolidation'. In: *2017 51st Asilomar Conference on Signals, Systems, and Computers*. 2017, pp. 336–343. DOI: 10.1109/ACSSC.2017.8335196.

[119] J A Behar et al. 'Sleep questionnaires in screening for obstructive sleep apnoea'. In: *2017 Computing in Cardiology (CinC)*. Sept. 2017, pp. 1–5. DOI: 10.22489/CinC.2017.233-188.

[120] B Kemp. 'The Sleep-EDF Database'. In: *World Wide Web, accessed August 2009* (2009).

[121] B Kemp. 'The Sleep-EDF Database [Expanded]'. In: *World Wide Web, accessed January 2015* (2015). URL: https://www.physionet.org/physiobank/database/sleep-edfx/.

[122] Christian O'reilly et al. 'Montreal Archive of Sleep Studies: An Open-access Resource for Instrument Benchmarking and Exploratory Research'. In: *Journal of Sleep Research* 23.6 (2014), pp. 628–635.

[123] Conor Heneghan. *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*.

[124] Sleep Heart Health Research Group. 'Methods for obtaining and analyzing unattended polysomnography data for a multicenter study'. In: *Sleep* 21.7 (1998), pp. 759–767.

[125] Stuart F Quan et al. ‘The sleep heart health study: design, rationale, and methods’. In: *Sleep* 20.12 (1997), pp. 1077–1085.

[126] Dennis A Dean et al. ‘Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource’. In: *Sleep* 39.5 (2016), pp. 1151–1164.

[127] Mario Giovanni Terzano et al. ‘Atlas, Rules, and Recording Techniques for the Scoring of Cyclic Alternating Pattern (CAP) in Human Sleep’. In: *Sleep Medicine* 2.6 (2001), pp. 537–553. ISSN: 13899457. DOI: 10.1016/S1389-9457(01)00149-6.

[128] George B Moody and Roger G Mark. ‘The impact of the MIT-BIH arrhythmia database’. In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 45–50.

[129] George B Moody and Roger G Mark. ‘The MIT-BIH arrhythmia database on CD-ROM and software for use with it’. In: *Computers in Cardiology 1990, Proceedings.* IEEE. 1990, pp. 185–188.

[130] S Devuyst, D Dutoit and M Kerkhofs. ‘The DREAMS Databases’. In: *The DREAMS Sleep Spindles Database. doi* 10 (2011).

[131] K Aggarwal et al. ‘A Structured Learning Approach with Neural Conditional Random Fields for Sleep Staging’. In: *2018 IEEE International Conference on Big Data (Big Data)*. Dec. 2018, pp. 1318–1327. DOI: 10.1109/BigData.2018.8622286.

[132] Y Yuan et al. ‘Multivariate Sleep Stage Classification using Hybrid Self-Attentive Deep Learning Networks’. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Dec. 2018, pp. 963–968. DOI: 10.1109/BIBM.2018.8621146.

[133] Kate McCarthy, Bibi Zabar and Gary Weiss. ‘Does cost-sensitive learning beat sampling for classifying rare classes?’ In: *Proceedings of the 1st international workshop on Utility-based data mining*. ACM. 2005, pp. 69–77.

[134] Gary M Weiss, Kate McCarthy and Bibi Zabar. ‘Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?’ In: *DMIN* 7 (2007), pp. 35–41.

[135] I N Yulita, M I Fanany and A M Arymurthy. 'Fuzzy Clustering and Bidirectional Long Short-Term Memory for Sleep Stages Classification'. In: *2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT)*. 2017, pp. 11–16. DOI: 10.1109/ICSIIT.2017.44.

[136] Edward Bedrosian. 'A product theorem for Hilbert transforms'. In: (1962).

[137] Ashish Vaswani et al. 'Attention is all you need'. In: *Advances in Neural Information Processing Systems*. 2017, pp. 5998–6008.

[138] James Justice. 'Analytic signal processing in music computation'. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27.6 (1979), pp. 670–684.

[139] Elie Aljalbout et al. 'Clustering with Deep Learning: Taxonomy and New Methods'. In: (2018).

[140] Jimmy Lei Ba, Jamie Ryan Kiros and Geoffrey E Hinton. 'Layer normalization'. In: *arXiv preprint arXiv:1607.06450* (2016).

[141] Guido F Montufar et al. 'On the number of linear regions of deep neural networks'. In: *Advances in neural information processing systems*. 2014, pp. 2924–2932.

[142] Kaiming He et al. 'Deep residual learning for image recognition'. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[143] Saining Xie et al. 'Aggregated residual transformations for deep neural networks'. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 5987–5995.

[144] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton. 'Imagenet classification with deep convolutional neural networks'. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[145] Min Lin, Qiang Chen and Shuicheng Yan. 'Network In Network'. In: *CoRR* abs/1312.4 (2013). URL: https://api.semanticscholar.org/CorpusID:16636683.

[146] David R So, Chen Liang and Quoc V Le. 'The Evolved Transformer'. In: *International Conference on Machine Learning*. 2019. URL: https://api.semanticscholar.org/CorpusID:59523610.

[147] Orestis Tsinalis, Paul M Matthews and Yike Guo. 'Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders'. In: *Annals of biomedical engineering* 44.5 (2016), pp. 1587–1597.

[148] Chris Drummond, Robert C Holte et al. 'C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling'. In: *Workshop on learning from imbalanced datasets II*. Vol. 11. Citeseer. 2003, pp. 1–8.

[149] Rich Caruana. 'Multitask learning'. In: *Machine learning* 28.1 (1997), pp. 41–75.

[150] Nasser M Nasrabadi. 'Pattern recognition and machine learning'. In: *Journal of electronic imaging* 16.4 (2007), p. 49901.

[151] Hao Dong et al. 'Mixed neural network approach for temporal sleep stage classification'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.2 (2018), pp. 324–333.

[152] Jacob Cohen. 'A Coefficient of Agreement for Nominal Scales'. In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.

[153] Kaiming He et al. 'Delving deep into rectifiers: Surpassing human-level performance on imagenet classification'. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.

[154] A I Humayun et al. 'End-to-end Sleep Staging with Raw Single Channel EEG using Deep Residual ConvNets'. In: *IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. May 2019, pp. 1–5. DOI: 10.1109/BHI.2019.8834483.

[155] Mart. 'TensorFlow: a system for large-scale machine learning'. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI'16. USA: USENIX Association, 2016, pp. 265–283. ISBN: 9781931971331.

[156] Maurice M Ohayon. 'Epidemiology of insomnia: what we know and what we still need to learn'. In: *Sleep Medicine Reviews* 6.2 (2002), pp. 97–111.

[157] Chiranth Bhagavan et al. 'Cannabinoids in the Treatment of Insomnia Disorder: A Systematic Review and Meta-Analysis'. In: *CNS Drugs* 34.12 (2020), pp. 1217–1228. ISSN: 1179-1934. DOI: 10.1007/s40263-020-00773-x.

[158] Ye Zhang et al. 'Worldwide and regional prevalence rates of co-occurrence of insomnia and insomnia symptoms with obstructive sleep apnea: A systematic review and meta-analysis'. In: *Sleep Medicine Reviews* 45 (2019), pp. 1–17. DOI: 10.1016/j.smrv.2019.01.004.

[159] Fifth Edition et al. 'Diagnostic and statistical manual of mental disorders'. In: *Am Psychiatric Assoc* (2013).

[160] A Crivello et al. 'The Meaning of Sleep Quality: A Survey of Available Technologies'. In: *IEEE Access* 7 (2019), pp. 167374–167390. DOI: 10.1109/ACCESS.2019.2953835.

[161] Hongle wu et al. 'Personal sleep pattern visualization using sequence-based kernel self-organizing map on sound data'. In: *Artificial Intelligence in Medicine* 80 (2017).

[162] Xinyu Huang et al. 'Sleep stage classification for child patients using DeConvolutional Neural Network'. In: *Artificial Intelligence in Medicine* 110 (2020), p. 101981. ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2020.101981. URL: https://www.sciencedirect.com/science/article/pii/S093336572031246X.

[163] Xi Jian Dai et al. 'Decreased modulation of segregated SEEKING and selective attention systems in chronic insomnia'. In: *Brain Imaging and Behavior* 15.1 (2021), pp. 430–443. ISSN: 19317565. DOI: 10.1007/s11682-020-00271-0.

[164] Sarah E Hogan et al. 'Slow-oscillation activity is reduced and high frequency activity is elevated in older adults with insomnia'. In: *Journal of Clinical Sleep Medicine* 16.9 (2020), pp. 1445–1454.

[165] Thierry Provencher, Shirley Fecteau and Célyne H Bastien. 'Patterns of Intrahemispheric EEG Asymmetry in Insomnia Sufferers: An Exploratory Study'. In: *Brain Sciences* 10.12 (2020), p. 1014. ISSN: 2076-3425. DOI: 10.3390/brainsci10121014.

[166] María Corsi-Cabrera, Olga Rojas-Ramos and Yolanda Río-Portilla. 'Waking EEG signs of non-restoring sleep in primary insomnia patients'. In: *Clinical Neurophysiology* 127 (2015), pp. 1813–1821. DOI: 10.1016/j.clinph.2015.08.023.

[167] Michele A Colombo et al. 'More severe insomnia complaints in people with stronger long-range temporal correlations in wake resting-state EEG'. In: *Frontiers in Physiology* 7 (2016), p. 576.

[168] Moi Hoon Yap et al. 'Breast ultrasound region of interest detection and lesion localisation'. In: *Artificial intelligence in medicine* 107 (2020), p. 101880. ISSN: 0933-3657.

[169] Ramin Nateghi, Habibollah Danyali and Mohammad Sadegh Helfroush. 'A deep learning approach for mitosis detection: Application in tumor proliferation prediction from whole slide images'. In: *Artificial intelligence in medicine* 114 (2021), p. 102048. ISSN: 0933-3657.

[170] Jose Bernal et al. 'Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review'. In: *Artificial intelligence in medicine* 95 (2019), pp. 64–81. ISSN: 0933-3657.

[171] C Dissanyaka et al. 'Classification of healthy and insomnia subjects based on wake-to-sleep transition'. In: *IEEE EMBS Conference on Biomedical Engineering and Sciences*. 2016, pp. 480–483.

[172] Serap Aydin, Hamdi Saraoğlu and Kara S. 'Singular spectrum analysis of sleep EEG in insomnia'. In: *Journal of Medical Systems* 35 (2011), pp. 457–461. DOI: 10.1007/s10916-009-9381-7.

[173] Wouter M Kouw and Marco Loog. 'An introduction to domain adaptation and transfer learning'. In: *arXiv preprint arXiv:1812.11806* (2018).

[174] Charles M Morin. *Insomnia: psychological assessment and management.* Guilford press, 1993.

[175] Charles M Morin et al. 'Behavioral and pharmacological therapies for late-life insomnia: a randomized controlled trial'. In: *The Journal of the American Medical Association* 281.11 (1999), pp. 991–999.

[176] Célyne H Bastien, Annie Vallières and Charles M Morin. 'Validation of the insomnia severity index as an outcome measure for insomnia research'. In: *Sleep Medicine* 2.4 (2001), pp. 297–307. ISSN: 1389-9457. DOI: https://doi.org/10.1016/S1389-9457(00)00065-4.

[177] I Chouvarda et al. ‘CAP sleep in insomnia: new methodological aspects for sleep microstructure analysis’. In: *IEEE Conference of Engineering in Medicine and Biology Society*. 2011, pp. 1495–1498.

[178] I Chouvarda et al. ‘Cyclic alternating patterns in normal sleep and insomnia: structure and content differences’. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20.5 (2012), pp. 642–652.

[179] Manish Sharma, Virendra Patel and U Rajendra Acharya. ‘Automated identification of insomnia using optimal bi-orthogonal wavelet transform technique with single-channel EEG signals’. In: *Knowledge-Based Systems* 224 (2021), p. 107078. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2021.107078.

[180] S T Hamida, T Penzel and B Ahmed. ‘EEG time and frequency domain analyses of primary insomnia’. In: *IEEE Conference of Engineering in Medicine and Biology Society*. 2015, pp. 6206–6209.

[181] S T Hamida, B Ahmed and T Penzel. ‘A novel insomnia identification method based on Hjorth parameters’. In: *IEEE International Symposium on Signal Processing and Information Technology*. 2015, pp. 548–552.

[182] S T Hamida et al. ‘How many sleep stages do we need for an efficient automatic insomnia diagnosis?’ In: *IEEE Conference of Engineering in Medicine and Biology Society*. 2016, pp. 2431–2434.

[183] H Yu et al. ‘Insomnia prediction using temporal feature of spindles’. In: *IEEE Conference on Healthcare Informatics*. 2019, pp. 1–8.

[184] Mostafa Shahin et al. ‘A two stage approach for the automatic detection of insomnia’. In: *IEEE Conference of Engineering in Medicine and Biology Society*. 2018, pp. 466–469. ISBN: 9781538636466. DOI: 10.1109/EMBC.2018.8512360.

[185] B Yang and H Liu. ‘Automatic identification of insomnia based on single-channel EEG labelled with sleep stage annotations’. In: *IEEE Access* 8 (2020), pp. 104281–104291.

[186] Q Wei et al. ‘A Residual based Attention Model for EEG based Sleep Staging’. In: *IEEE Journal of Biomedical and Health Informatics* 24 (2020), pp. 2833–2843. DOI: 10.1109/JBHI.2020.2978004.

[187] Yaroslav Ganin et al. 'Domain-adversarial training of neural networks'. In: *The Journal of Machine Learning Research* 17.1 (2016), pp. 2030–2096.

[188] Mathieu Salzmann et al. 'Factorized Orthogonal Latent Spaces'. In: *International Conference on Artificial Intelligence and Statistics*. Vol. 9. 2010, pp. 701–708.

[189] Hasim Sak, Andrew W Senior and Françoise Beaufays. 'Long short-term memory recurrent neural network architectures for large scale acoustic modeling'. In: *Conference of the International Speech Communication Association*. 2014, pp. 338–342.

[190] Terrence D Lagerlund. 'Manipulating the Magic of Digital EEG: Montage Reformatting and Filtering'. In: *American Journal of Electroneurodiagnostic Technology* 40.2 (2000), pp. 121–136. DOI: 10.1080/1086508X.2000.11079295. URL: https://doi.org/10.1080/1086508X.2000.11079295.

[191] Mohammad Rezaei, Hiwa Mohammadi and Habibolah Khazaie. 'EEG/EOG/EMG data from a cross sectional study on psychophysiological insomnia and normal sleep subjects'. In: *Data in Brief* 15 (Sept. 2017), pp. 314–319. ISSN: 23523409. DOI: 10.1016/j.dib.2017.09.033. URL: https://data.mendeley.com/datasets/3hx58k232n/4.

[192] Sangheum Hwang and Hyo-Eun Kim. 'Self-transfer learning for weakly supervised lesion localization'. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016, pp. 239–246.

[193] Martin Reite et al. 'The use of polysomnography in the evaluation of insomnia'. In: *Sleep* 18.1 (1995), pp. 58–70.

[194] Simon Hartmann et al. 'Characterization of Cyclic Alternating Pattern during Sleep in Older Men and Women using Large Population Studies'. In: *Sleep* 43.7 (May 2020). ISSN: 15509109. DOI: 10.1093/SLEEP/ZSAA016.

[195] Mario Giovanni Terzano and Liborio Parrino. 'Origin and Significance of the Cyclic Alternating Pattern (CAP): REVIEW ARTICLE'. In: *Sleep Medicine Reviews* 4.1 (2000), pp. 101–123. ISSN: 1087-0792. DOI: https://doi.org/10.1053/smrv.1999.0083.

[196] Manish Sharma et al. 'CAPSCNet: A Novel Scattering Network for Automated Identification of Phasic Cyclic Alternating Patterns of Human Sleep using Multivariate

EEG Signals'. In: *Computers in Biology and Medicine* 164 (2023), p. 107259. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2023.107259.

[197] Umberto Barcaro et al. 'A General Automatic Method for the Analysis of NREM Sleep Microstructure'. In: *Sleep Medicine* 5.6 (2004), pp. 567–576. ISSN: 1389-9457. DOI: https://doi.org/10.1016/j.sleep.2004.07.012.

[198] Manish Sharma et al. 'Computerized Detection of Cyclic Alternating Patterns of Sleep: A New Paradigm, Future Scope and Challenges'. In: *Computer Methods and Programs in Biomedicine* 235 (2023), p. 107471. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2023.107471.

[199] Hogeon Seo et al. 'Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG'. In: *Biomedical Signal Processing and Control* 61 (2020), p. 102037. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2020.102037.

[200] Sajad Mousavi, Fatemeh Afghah and U Rajendra Acharya. 'SleepEEGNet: Automated Sleep Stage Scoring with Sequence to Sequence Deep Learning Approach'. In: *PLOS ONE* 14.5 (2019), e0216456. DOI: 10.1371/journal.pone.0216456.

[201] Huy Phan et al. 'SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging'. In: 27 (2018), pp. 440–410.

[202] Akara Supratak and Yike Guo. 'TinySleepNet: An Efficient Deep Learning Model for Sleep Stage Scoring based on Raw Single-Channel EEG'. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020, pp. 641–644. DOI: 10.1109/EMBC44109.2020.9176741.

[203] Jianan Ye et al. 'CoSleep: A Multi-View Representation Learning Framework for Self-Supervised Learning of Sleep Stage Classification'. In: *IEEE Signal Processing Letters* 29 (2022), pp. 189–193. DOI: 10.1109/LSP.2021.3130826.

[204] H Phan et al. 'XSleepNet: Multi-View Sequential Model for Automatic Sleep Staging'. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 44.09 (Sept. 2022), pp. 5903–5915. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2021.3070057.

[205] Yujie Li et al. ‘MVF-SleepNet: Multi-View Fusion Network for Sleep Stage Classification’. In: *IEEE Journal of Biomedical and Health Informatics* 28.5 (2024), pp. 2485–2495. DOI: 10.1109/JBHI.2022.3208314.

[206] Mingsheng Long et al. ‘Learning Transferable Features with Deep Adaptation Networks’. In: *Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37. PMLR, May 2015, pp. 97–105.

[207] Y Balaji, R Chellappa and S Feizi. ‘Normalized Wasserstein for Mixture Distributions With Applications in Adversarial Learning and Domain Adaptation’. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, 2019.

[208] Fuzhen Zhuang et al. ‘Supervised Representation Learning: Transfer Learning with Deep Autoencoders’. In: *International Joint Conference on Artificial Intelligence*. 2015.

[209] Kuniaki Saito et al. ‘Maximum Classifier Discrepancy for Unsupervised Domain Adaptation’. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2017), pp. 3723–3732.

[210] X Peng et al. ‘Moment Matching for Multi-Source Domain Adaptation’. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, 2019.

[211] Hang Wang et al. ‘Learning to Combine: Knowledge Aggregation for Multi-source Domain Adaptation’. In: *The European Conference on Computer Vision (ECCV)*. 2020.

[212] Luyu Yang et al. ‘Curriculum Manager for Source Selection in Multi-Source Domain Adaptation’. In: *The European Conference on Computer Vision (ECCV)*. 2020.

[213] Zhongying Deng et al. ‘Dynamic Instance Domain Adaptation’. In: *IEEE Transactions on Image Processing* 31 (2022), pp. 4585–4597. DOI: 10.1109/TIP.2022.3186531.

[214] Artem Rozantsev, Mathieu Salzmann and Pascal Fua. ‘Beyond Sharing Weights for Deep Domain Adaptation’. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.4 (2019), pp. 801–814. DOI: 10.1109/TPAMI.2018.2814042.

[215] Róger Bermúdez Chacón, Mathieu Salzmann and Pascal Fua. 'Domain-adaptive Multibranch Networks'. In: *International Conference on Learning Representations (ICLR)*. 2020.

[216] Kihyuk Sohn et al. 'FixMatch: Simplifying Semi-supervised Learning with Consistency and Confidence'. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020.

[217] Huy Phan et al. 'Towards More Accurate Automatic Sleep Staging via Deep Transfer Learning'. In: *IEEE Transactions on Biomedical Engineering* 68.6 (2021), pp. 1787–1798. DOI: 10.1109/TBME.2020.3020381.

[218] Songlu Lin et al. 'SSC-SleepNet: A Siamese-Based Automatic Sleep Staging Model with Improved N1 Sleep Detection'. In: *IEEE Journal of Biomedical and Health Informatics* (2025), pp. 1–13. DOI: 10.1109/JBHI.2025.3572886.

[219] Leyuan Huang et al. 'EEGCMCNet: A Hybrid Network for Sleep Stage Classification'. In: *IEEE Sensors Journal* (2025), p. 1. DOI: 10.1109/JSEN.2025.3574141.

[220] S M Nuruzzaman Nobel et al. 'CRT: A Convolutional Recurrent Transformer for Automatic Sleep State Detection'. In: *IEEE Journal of Biomedical and Health Informatics* 29.6 (2025), pp. 4452–4462. DOI: 10.1109/JBHI.2025.3543028.

[221] Zhen Yu et al. 'IR-Based Sleep Monitoring: Movement, Respiration, and Sleep Staging With a Structure-Aware Cross-Modal EEG Knowledge Distillation'. In: *IEEE Internet of Things Journal* (2025), p. 1. DOI: 10.1109/JIOT.2025.3583049.

[222] Ze Ren, Jin Ma and Ying Ding. 'FlexibleSleepNet:A Model for Automatic Sleep Stage Classification Based on Multi-Channel Polysomnography'. In: *IEEE Journal of Biomedical and Health Informatics* 29.5 (2025), pp. 3488–3501. DOI: 10.1109/JBHI.2025.3525626.

[223] Ruhan Liu et al. 'MHFNet: A Multimodal Hybrid-Embedding Fusion Network for Automatic Sleep Staging'. In: *IEEE Journal of Biomedical and Health Informatics* 29.5 (2025), pp. 3387–3397. DOI: 10.1109/JBHI.2025.3528444.

[224] Dongdong Zhou et al. 'Interpretable Sleep Stage Classification Based on Layer-Wise Relevance Propagation'. In: *IEEE Transactions on Instrumentation and Measurement* 73 (2024), pp. 1–10. DOI: 10.1109/TIM.2024.3370799.

[225] Yuxuan Yang and Yanli Li. 'A Lightweight Model for Physiological Signals-Based Sleep Staging With Multiclass CAM for Model Explainability'. In: *IEEE Sensors Journal* 24.17 (2024), pp. 27815–27823. DOI: 10.1109/JSEN.2024.3430009.

[226] Changqing Lu et al. 'Channel Contribution in Deep Learning Based Automatic Sleep Scoring—How Many Channels Do We Need?' In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), pp. 494–505. DOI: 10.1109/TNSRE.2022.3227040.

[227] Jonghyun Park et al. 'Ultra-Wideband Radar-Based Sleep Stage Classification in Smartphone Using an End-to-End Deep Learning'. In: *IEEE Access* 12 (2024), pp. 61252–61264. DOI: 10.1109/ACCESS.2024.3390391.

[228] Qiongyan Wang, Hanrong Cheng and Wenjin Wang. 'Video-PSG: An Intelligent Contactless Monitoring System for Sleep Staging'. In: *IEEE Transactions on Biomedical Engineering* 72.3 (2025), pp. 965–977. DOI: 10.1109/TBME.2024.3480813.

[229] Bing Zhai et al. 'DSleepNet: Disentanglement Learning for Personal Attribute-Agnostic Three-Stage Sleep Classification Using Wearable Sensing Data'. In: *IEEE Journal of Biomedical and Health Informatics* 29.7 (2025), pp. 4861–4872. DOI: 10.1109/JBHI.2025.3541851.