

INTERPRETABLE MULTI-MODAL SLEEP MONITORING SYSTEM USING EAR-EEG AND EOG

Dhinesh S (170133B)

Jathurshan P (170248G)

Mithunjha A (170389M)

Vinith K (170654X)

Bachelor of Science of Engineering in Biomedical Engineering

The Department of Electronic & Telecommunication Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2022

**INTERPRETABLE MULTI-MODAL SLEEP
MONITORING SYSTEM USING
EAR-EEG AND EOG**

Dhinesh S (170133B)

Jathurshan P (170248G)

Mithunjha A (170389M)

Vinith K (170654X)

Thesis submitted in partial fulfillment of the requirements for the degree
Bachelor of Science of Engineering in Biomedical Engineering

The Department of Electronic & Telecommunication Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2022

Signature:

Date:

Signature:

Date:

Signature:

Date:

Signature:

Date:

The supervisors should certify the Thesis with the following declaration.

The above candidate has carried out research for the Bachelor of Science of Engineering in Biomedical Engineering Thesis under our supervision. We confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Anjula De Silva

Signature of the Supervisor:

Date:

Name of Supervisor: Dr. Chamira Edusooriya

Signature of the Supervisor:

Date:

Name of Supervisor: Dr. Simon Lind Kappel

Signature of the Supervisor:

Date:

APPROVAL

To our Parents, Lecturers, Friends
and
the People of the Democratic Socialist Republic of Sri Lanka

DECLARATION

To our Parents, Lecturers, Friends
and
the People of the Democratic Socialist Republic of Sri Lanka

ABSTRACT

Polysomnography is the gold standard test used for sleep disorder diagnosis in which multiple bioelectric signals are acquired followed by manual annotation of sleep stages by sleep technicians. The conventional sleep study approach requires manual annotation and multiple signals which in return affects the reliability and comfort of the process. Additionally, deep learning has been explored in the past for automatic sleep stage classification but their black-box behavior and lack of interpretability keeps them from being used in clinical environments. The focus of this final year project is to develop a sleep monitoring system with automatic sleep stage classification using deep-learning with interpretability component and enhanced level of comfort. This is achieved by designing an ear-oriented signal acquisition system that acquires ear-EEG and EOG modalities. In past there have been several efforts to build comfortable home based sleep monitoring system using ear-EEG. But the performance was inferior compared to scalp EEG based sleep staging. The project also focuses on leveraging deep learning algorithm to improve ear-EEG based sleep staging

Objectives : Designing a user-comfortable hardware setup with ear-EEG and EOG modalities for sleep stage monitoring; Development of state-of-the-art sleep stage classification algorithm with interpretability aspect using EEG and EOG modalities; Adapting the sleep stage classification algorithm towards ear-EEG to enhance performance; Development of accurate sleep stage classification algorithm using ear-EEG and EOG signals.

Algorithm Development : Accurate sleep stage classification is significant for sleep health assessment. Recently, deep learning methods have been utilized to obtain improved accuracy on sleep stage classification algorithms. However, the sleep experts are skepticism in using deep learning for sleep staging due to their black-box behaviour. We propose *Cross-Modal Transformers* to achieve competitive performance with the current state-of-the-art methods and have the interpretability component. We employed a well-known method named *cross-domain knowledge distillation* with cross-modal transformers to to improve the ear-EEG based sleep staging performance. The final algorithm was developed using ear-EEG and EOG signals based on cross-modal transformers.

Hardware Development : Maintaining the sleep quality plays an important role in sleep stage classification and sleep health assessment. Polysomnography studies disrupts comfort of the user and thereby disturbs the quality of sleep. Therefore, developing a compact and user friendly hardware for the collection of EEG and EOG signals is of great importance for maintaining the quality of sleep. A compact ear piece that incorporates ear-EEG sensors with an external EOG sensor is the proposed design for the hardware implementation.

Keywords: Ear-EEG, Sleep stage classification, Transformers, Interpretable models, Cross-modal knowledge distillation

DEDICATION

To our Parents, Lecturers, Friends
and
the People of the Democratic Socialist Republic of Sri Lanka

ACKNOWLEDGEMENT

We would like to extend our gratitude to the supervisors of our project, Dr. Anjula De Silva and Dr. Chamira Edussooriya, and external collaborator of the project, Dr. Simon Lind Kappel for their immense support, guidance and encouragement.

We extend our gratitude towards Wickramarachchi Institute of Speech and Hearing (WISH) for their support in acquiring the ear-molds for our research, Dr. Ranga Rodrigo and the National Research Council of Sri Lanka for providing computational resources for conducting our experiments.

We also extend our gratitude for the family members of the group members for the support given to make this project a success.

Our gratitude also extended to all the academic and non-academic staff of Department of Electronic and Telecommunication, University of Moratuwa, as well as our fellow batchmates for the support provided.

TABLE OF CONTENTS

Approval	ii
Declaration	iii
Abstract	v
Dedication	vi
Acknowledgement	vii
Table of Contents	viii
List of Figures	xi
List of Tables	xiii
Abbreviations	xiv
1 Introduction	1
1.1 Problem Statement	1
1.2 Main Objective	2
1.3 Project Scope	2
1.4 Proposed System Architecture	3
1.5 Novelty and the Uniqueness of the Project	4
1.6 Potential Applications	4
1.7 Organization of the Thesis	5
2 Literature Review	6
2.1 Scalp EEG for Sleep Stage Classification	6
2.2 Ear-EEG for Sleep Stage Classification	7
2.3 Other Physiological Signals for Sleep Stage Classification	7
2.4 Existing Hardware	9
3 Sleep Stage Classification Algorithm	11
3.1 Signal Processing and Feature Extraction	11
3.2 Base Model and Interpretable Deep Learning Algorithm Development for Sleep Stage Classification	13
3.2.1 Analysis of Alternatives	13

3.2.2	Epoch Transformer	14
3.2.3	Sequence Transformer	15
3.2.4	Cross-Modal Transformers	17
3.3	Adaptation of Algorithm Towards Ear-EEG	23
3.3.1	Dataset	23
3.3.2	Data preparation	23
3.3.3	Supervised Learning on Ear-EEG	25
3.3.4	Transfer Learning	25
3.3.5	Knowledge Distillation	26
3.3.6	Training Setup	27
3.3.7	Results	28
3.3.8	Final algorithm	29
4	Signal Acquisition Hardware Setup	31
4.1	Initial Study	31
4.2	Preliminary Design	31
4.3	Overview of the Hardware System	32
4.3.1	The Intended system	32
4.3.2	The Earpiece	33
4.3.3	The enclosure design	34
4.3.4	Signal acquisition device	34
4.4	Challenges in Execution	38
4.5	The Experimental procedure	39
4.6	Ethical Clearance	40
4.7	Future Prospects	40
5	Results and Discussion	42
5.1	Cross-Modal Transformer	42
5.1.1	Experiments	42
5.1.2	Sleep Stage Classification Performance	43
5.1.3	Computational Complexity	45
5.1.4	Interpretation results	46
5.2	Validation of Hardware Implementation	48

References	50
Appendices	55

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Illustration of discomfort faced by the patient during polysomnography test [5].	1
Figure 1.2	Overall architecture of the proposed sleep monitoring system	3
Figure 1.3	Overall architecture of the final sleep monitoring system	4
Figure 2.1	The chosen electrode placements	9
Figure 3.1	Overview of the sleep stage classification algorithm.	11
Figure 3.2	Short time Fourier transform analyses between different sleep stages on the sleep-EDF-expanded dataset.	12
Figure 3.3	Power spectral density analyses between different sleep stages on the sleep-EDF-expanded dataset.	12
Figure 3.4	Comparison between deep learning architectures.	13
Figure 3.5	Architecture of the epoch transformer experimented for sleep stage classification on sleep-EDF-expanded 2018 dataset.	14
Figure 3.6	Architecture of the sequence transformer experimented for sleep stage classification on sleep-EDF-expanded 2018 dataset.	16
Figure 3.7	One-to-one and many-to-many classification schemes.	18
Figure 3.8	Architecture of proposed epoch cross-modal transformer	20
Figure 3.9	Architecture of proposed sequence cross-modal transformer	21
Figure 3.10	Method for interpretation	22
Figure 3.11	The correspondence between scalp and ear EEG derivations. Left - ear EEG derivations (L-R, L, R) and Right - scalp ear EEG derivations (A1-A2, C3-O1, C4-O2)	24
Figure 3.12	Supervised Training Framework	25
Figure 3.13	Transfer Learning Framework	26
Figure 3.14	Knowledge distillation Framework	26
Figure 3.15	Online Knowledge distillation Framework	27
Figure 3.16	Offline Knowledge distillation frame work	28
Figure 3.17	Confusion matrix of ear EEG only (left) and ear EEG and EOG signal (right)	30
Figure 3.18	Interpretation with ear EEG and EOG signals	30
Figure 4.1	The overall block diagram of the intended ear-EEG acquisition device	32
Figure 4.2	Types of ear molds	33
Figure 4.3	The CAD and printed enclosure design	34
Figure 4.4	The enclosure design inspired from Muse 2 bands	34

Figure 4.5	Modules used to build the modular level signal acquisition device	35
Figure 4.6	The Power Supply	36
Figure 4.7	This schematics shows the electrode input terminals, the active shielding and the passive low pass filtering blocks	36
Figure 4.8	ADS1299 - The Analog Front End Circuit	37
Figure 4.9	STM32 - The Micro-controller Circuit	38
Figure 4.10	Basic study procedures conducted on the subject	39
Figure 4.11	Sleep study procedures conducted on the subject	40
Figure 5.1	Interpretation results for sequence of PSG epochs in N3, N3, N3, N2 and wake stages respectively.	46
Figure 5.2	Interpretation results for sequence of PSG epochs in REM, REM, wake, wake and N1 stages respectively.	47
Figure 5.3	ECG artifacts observed during ear EEG signal acquisition	48
Figure 5.4	Power Spectral Density plot of the ear EEG data acquired during the sleep study	49
Figure 5.5	STFT magnitude plots of alpha, beta, theta and delta frequency bands with time, for the ear EEG data acquired during the sleep study	49

LIST OF TABLES

Table	Description	Page
Table 3.1	Handcrafted features extracted for feature importance analysis	12
Table 3.2	Performance of the epoch transformer on sleep-EDF-expanded 2018 dataset	15
Table 3.3	Performance of the sequence transformer on sleep-EDF-expanded 2018 dataset	15
Table 3.4	Description of the rejected channels	24
Table 3.5	Performance comparison of Epoch Cross-Modal Transformer on ear-EEG Dataset with state-of-the-art	25
Table 3.6	Results of Transfer Learning and Knowledge Distillation	28
Table 3.7	Performance comparison between supervised training and knowledge distillation	29
Table 3.8	Performance of Ear-EEG and EOG based sleep staging	29
Table 5.1	Performance comparison between cross-modal transformer and previous works on sleep-EDF-expanded 2018 dataset.	44
Table 5.2	Class-wise performance comparison between cross-modal transformer and previous works on sleep-EDF-expanded 2018 dataset.	45
Table 5.3	Comparison of model size and training time.	45

ABBREVIATIONS

AASM – American Academy of Sleep Medicine
ADC - analog to digital converter
Ag - Silver
AgCl - Silver Chloride
BLE - Bluetooth Low Energy
BME - Biomedical Engineering
BMI – Body Mass Index
CAD - Computer Aided Designing
CWT - Continuous Wavelet Transform
CinC - Computing in Cardiology
CNN – Convolutional Neural Network
DBN - Deep Belief Networks
DC - Direct current
ECG – Electrocardiogram
EEG - Electroencephalogram
EMG – Electromyogram
EOG – Electrooculogram
HRV - Heart Rate Variability
IR - Infrared
LOSO - Leave One Subject Out
LPF - low pass filter
LSTM – Long Short-Term Memory
MASS - Montreal Archive of Sleep Studies
MLP - Multi Layer Perceptron
MPG – Mechanical Plethysmography
NREM – Non-Rapid Eye Movement
PCB - Printed circuit board
PELV - Protective Extra Low Voltage
PPG - Photoplethysmogram
PSD - Power Spectral Density
PSG - Polysomnography
R&K - Rechtschaffen and Kales
RC - resistive/capacitive
REM – Rapid Eye Movement
RNN - recurrent neural network
RRV - Respiratory Rate Variability

SA – Sleep Apnea

SHHS - Sleep Heart Health Study

SNR - Signal to Noise Ratio

SOC - System On Chip

SPI - Serial Peripheral Interface

STFT - Short Time Fourier Transform

TPU - Thermoplastic Polyurethane

USB - Universal Serial Bus

Chapter 1

INTRODUCTION

Humans spend around one-third of their life-time sleeping, which is crucial to protect their mental and physical health. Sleep disorders are becoming a common health problem worldwide, due to stress, depression and insufficient sleep. A survey conducted in US between 1999 and 2004 shows that 50 – 70 million adults are affected over 70 different kinds of sleep disorders[1].

Polysomnography (PSG), i.e sleep study, is a comprehensive and the gold standard test used to diagnose sleep disorders. During a Polysomnography test, several physiological signals such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electromyogram (EMG), Electrooculogram (EOG) and SpO₂ are recorded, to aid sleep stage detection and disorder diagnosis.

Humans go into different stages of sleep during their sleep cycle and it consists of three main stages: Wake, Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM). NREM can be further divided into sub-stages. The accurate identification of sleep stages is a fundamental step in sleep assessment, where 30-second PSG epochs are analysed to determine their sleep stages. Sleep staging is conducted manually by sleep technicians based on guidelines. American Academy of Sleep Medicine (AASM) [2] and Rechtschaffen and Kales (R&K)[3, 4] are the two international guidelines for sleep study, where AASM standards is most commonly used.

1.1 Problem Statement

During sleep study, patients will be wearing multiple sensors and electrodes in their body as shown in Fig. 1.1. Quality of patients' sleep is compromised due to the discomfort caused by these sensor systems and unfamiliarity of the hospital environment, thus affects the diagnostic quality of the study, leading to treatment errors. After the



Fig. 1.1: Illustration of discomfort faced by the patient during polysomnography test [5].

sleep study, sleep experts will go through around 6hrs of data and manually annotates each sleep stages. The conventional sleep stage scoring approach is labour intensive, time consuming, prone to human errors and costly. These two critical issues in sleep studies explain the necessity for an automatic sleep stage classification algorithm and a comfortable signal acquisition system.

Despite the improve performance by deep-learning based automatic sleep stage classification algorithms, their black-box behavior and lack of interpretability keeps them from being using in clinical environments. The sleep experts are skepticism in using deep learning due to this behavior. This shows the need for an interpretable sleep stage classification algorithm. In past there have been several efforts to build comfortable home based sleep monitoring system using ear-EEG, but the performance was inferior compared to scalp-EEG based sleep staging. We identified a gap in the research towards developing algorithms to enhance the performance of ear-EEG based sleep staging.

1.2 Main Objective

In our research we will be addressing the four major issues in sleep studies, which are discomfort caused due to existing signal acquisition systems, manual sleep stage classification, black-box behaviour of deep-learning based sleep staging algorithms and low performance of ear-EEG based sleep staging. The main objective of our study is to design and develop a multi-modal sleep monitoring system for accurate sleep stage classification with enhanced comfort to the users. In order to achieve our goal we developed a transformer based automatic sleep stage classification algorithm with interpretability using ear-EEG and EOG as a solution to address the shortcomings of manual sleep stage classification and black-box behaviour of deep-learning models. We enhance the performance of ear-EEG based sleep staging by leveraging a scalp-EEG assisted knowledge distillation framework. In order to address the discomfort problem, we designed a comfortable ear-EEG and EOG signal acquisition system.

1.3 Project Scope

Scopes and objectives of the research were defined in order to achieve our goal. The two main objectives of our research are as follows:

- **Development of sleep stage classification algorithm:** The main objective is to develop a sleep stage classification algorithm using ear-EEG and EOG signals. The development of sleep stage classification algorithm can be divided into the following sub-objectives:
 - Signal processing and feature extraction.

- Base model and interpretable deep learning algorithm (Cross-Modal Transformers).
 - Development of a knowledge distillation framework for scalp-EEG assisted enhancement of ear-EEG based sleep staging.
 - Final sleep stage classification algorithm development using ear-EEG and EOG.
- **Design a hardware for signal acquisition system:** In the hardware design, our focus would be on designing a system to acquire ear-EEG and EOG signals, considering the comfortness to the user and quality of the acquired signals.

The deliverables of our research would be a device for ear-EEG and EOG signal acquisition and a novel deep learning based sleep stage classification algorithm. We are expecting to submit the findings of our research to well-known conferences.

1.4 Proposed System Architecture

The initial proposed architecture given in Fig. 1.2 gives an overall view of our proposed solution. Initially, ear-EEG and EOG signals will be acquired from the patient using our signal acquisition system. Then the acquired signals will be sent through our sleep stage classification algorithm for sleep stage prediction. Our algorithm will consist of separate signal processing pipelines for ear-EEG and EOG signals and will undergo a common feature extraction process. After feature extraction, sleep stage for each epoch of acquired signals will be predicted with the aid of a deep learning based algorithm.

Figure 1.3 illustrates the final architecture of the proposed sleep monitoring system, where in addition to sleep stage classification, interpretability component is added to the system.

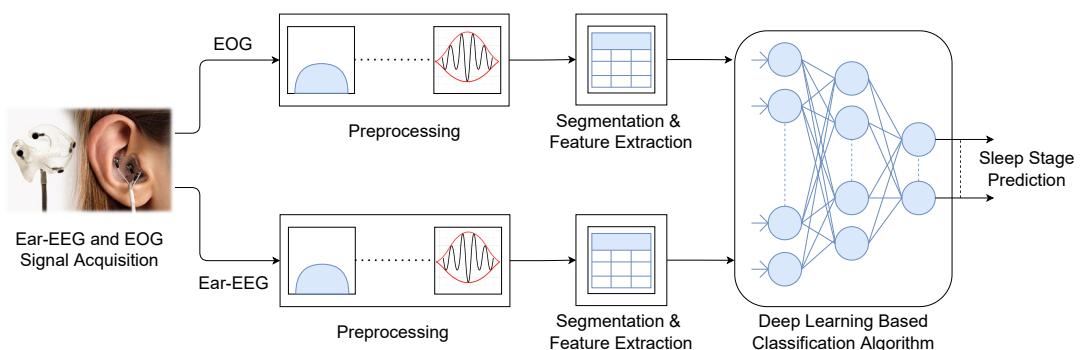


Fig. 1.2: Overall architecture of the proposed sleep monitoring system

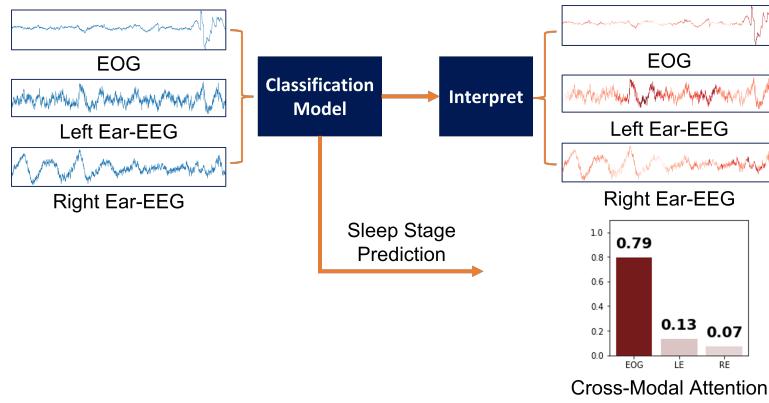


Fig. 1.3: Overall architecture of the final sleep monitoring system

1.5 Novelty and the Uniqueness of the Project

The main novelty of our research is the combination of ear-EEG with EOG for sleep monitoring. Ear-EEG based sleep monitoring is an emerging trend in research, but it hasn't been combined with EOG for sleep monitoring. The comfort to the user will be enhanced through our solution thus quality of sleep and quality of diagnosis will be improved. Our solution would be able classify sleep stages to aid sleep disorder diagnosis. In order to achieve that, classification algorithm will be initially developed using scalp EEG and EOG, and then will be adapted towards ear-EEG and EOG. The novelty of our sleep stage classification algorithm lies in its novel transformer architecture to learn intra-modal, cross-modal and inter epoch relationships and the ability to interpret the results. In addition, our proposed solution has the capability to address the growing need of home-based sleep monitoring.

1.6 Potential Applications

Main application of our research is home-based sleep monitoring system. Currently, sleep studies are conducted at sleep clinics due to the requirement of multiple signals. Our main focus of the research is to reduce the number of signals acquired while maintaining the level of accuracy at an acceptable range. This enables development of a home-friendly device to perform sleep monitoring as well as sleep stage detection. Also, since reduction in signals acquired increases the quality of patients' sleep, our research can also enhance the reliability and practicality of sleep monitoring at sleep clinics. In addition, our research also focuses on development of an accurate automatic sleep stage detection algorithm. Such automatic sleep staging algorithm can aid the sleep technicians and with sufficient accuracy it can even replace the sleep technicians as well.

1.7 Organization of the Thesis

This section focuses the navigation through our report. The objective, scope, overview of the proposed architecture, novelty and potential application of our final year project are covered in the Chapter 1. Chapter 2 focuses on the literature review and insights on related works. Chapter 3 and 4 focuses on the our algorithm development and hardware implementation, in order to achieve research objectives and their results are provided in Chapter 5.

Chapter 2

LITERATURE REVIEW

In order to develop an automatic sleep stage classification deep learning algorithm, approaches carried out in past literature were studied. To design a comfortable signal acquisition setup, different physiological signals (ear-EEG, EOG, ECG, ear-PPG) that can reflect sleep patterns are studied.

2.1 Scalp EEG for Sleep Stage Classification

In the last decade, scalp EEG signal was the most popular modality (60%) used in deep learning based automated sleep stage classification approaches [6]. Also, majority of the research have used deep learning-based approaches for automatic sleep stage classification [6]. Ability to learn and extract optimized features automatically from larger datasets is the main advantage of deep learning-based approaches over conventional feature-based approaches. Different types of deep learning-based architectures such as Convolutional Neural Networks (CNNs) [7–10], Recurrent Neural Networks (RNNs) [11, 12], Deep Belief Networks (DBNs) [13], Autoencoders [14] and transformers [15] have been utilized for automatic sleep stage classification. Hybrid architectures such as CNN with RNN [16, 17] and Deep Neural Networks with RNN [18] have also been utilized. We believe that the performance of CNN has saturated in this domain [6]. Also, the sleep stages are sequentially connected and cyclically alternated during sleep, but only a few research have utilized the sequential information. Transformers are the emerging trend in deep learning, which has shown its potential in various applications such as natural language processing, computer vision, etc. Current studies are also focusing on adapting transformer towards medical applications to improve the interpretability of the models [15].

Previous works have used sleep study datasets such as SleepEDF, sleep-EDF-expanded, SHHS (Sleep Heart Health Study), MIT-BIH and MASS to validate sleep stage classification algorithms. CNNs with EEG were able to achieve the accuracy of 85% in sleep-EDF-expanded dataset and overall, their performance has been in the range of 80-87%. CNNs were able to achieve 86.5% in MASS dataset and RNNs were able to achieve the accuracy of 85.9%. Hybrid architecture with CNN and RNN with EEG was able to achieve accuracy of 86.2% in the same dataset.

Based on the results, we can observe that scalp EEG based approaches yield high classification accuracy, but the comfort to the user is questionable. A significant number of previous works have focused on utilizing the mixture of other signals within the

PSG recordings such as EOG and EMG. Even though the ECG HRV is an important sleep parameter, only a few research have used it for sleep stage classification.

2.2 Ear-EEG for Sleep Stage Classification

Scalp EEG based sleep stage classification can yield high accuracy which is comparable to the manual annotations of sleep technicians, but the subjects' comfort level is questionable with scalp EEG and it is unsuitable for long-term sleep monitoring. This led to the investigation into light-weight and comfortable EEG alternatives, i.e., ear-EEG based approaches.

Ear-EEG approaches certainly mean that electrodes are placed inside or around the ears (outer ear). Currently, ear-EEG based sleep stage classification is an emerging trend in sleep study. Most of the previous works in this area have focused on designing signal acquisition systems and validating that ear-EEG can replace scalp EEG in sleep study. In previous research, different types of ear-EEG acquisition hardware setups (in ear or outer ear) have been developed and at the same time a good correspondence between ear-EEG and scalp EEG has been proven.

However, there is a high discrepancy between accuracy and Cohen's kappa (coefficient of agreement) attained by the scalp EEG based approaches and ear-EEG based approaches. Nakamura et al., 2017 [19] reported 4 class classification (excluding REM stage) accuracy of 76.8% and kappa of 0.65 with ear-EEG signal. Nakamura et al., 2019 [20] reported 5 class classification accuracy of 85.9% and kappa of 0.79 with scalp EEG signal but accuracy of 74.1% and kappa of 0.61 with ear-EEG signal. Mikkelsen et al., 2019 [21] reported an average kappa of 0.73 with ear-EEG signal.

Unfortunately, there are no publicly available datasets for ear-EEG. Most research studies collected their own ear-EEG dataset to conduct their study. As mentioned earlier, most of the previous work based on ear-EEG are mainly focused on validating the fact that ear-EEG can efficiently replace the scalp EEG in sleep study. Simple machine learning models like random forest classifier [21] and multi-class support vector machine with radial based function kernel [19] were employed to perform automatic sleep stage classification using ear-EEG. To our best knowledge deep learning based architectures has not been utilized for ear-EEG based sleep stage classification. Ear-EEG based sleep study area is not fully explored yet, and this could be due to the unavailability of public dataset.

2.3 Other Physiological Signals for Sleep Stage Classification

Although EEG has been the most successful modality for sleep stage classification, it has suffered from complexity and discomfort in the signal acquisition. On the

other hand, other physiological signals such as electrocardiography, photoplethysmography and respiratory signals (commonly termed as cardio-respiratory signals), electrooculography, electromyography and in-ear pressure [22] can be obtained with user-friendly devices, but the accuracy of sleep stage classification using them in the past were hardly in the clinicians-accepted range [23]. Also, past research on sleep staging with other physiological signals that obtained good results with healthy subjects were not able to achieve similar performance on patients, thus lacked robustness. To bridge this gap, studies have been conducted in recent times on deep learning-based sleep stage classification using cardio-respiratory signals, especially to address the issue of comfort. Other physiological signals have less information regarding sleep stages compared to EEG. Thus, the current studies on this area are trying to maximize utilization of the available information through deep learning.

EOG signals reflect the activities of eye, and in the context of sleep stage classification eye movements is a key indicator in distinguishing REM and NREM stages. Also, EOG electrode placements are closer to the EEG acquisition area, especially to Fp1 and Fp2 [24]. Thus, EOG recordings are also partially influenced by EEG signals, which can be utilized for sleep stage classification. Also, sleep stages are connected with activity of autonomous nervous system, i.e., REM sleep corresponds with sympathetic activity whereas NREM sleep corresponds to parasympathetic activity. Peak to peak intervals, which are also termed as R-R intervals in heart related signals such as ECG and PPG are good indicators of sympathetic and parasympathetic activity. These RR derived features are also commonly known as Heart Rate Variability (HRV) metrics. Past studies have established the correlation between sleep stages and HRV metrics and thus, physiological signals such as ECG and PPG can be utilized for sleep stage classification [24]. Similar to HRV, Respiratory Rate Variability (RRV) is derived from the respiratory signals and also serve as an indicator of sleep staging. However, both ECG and PPG signals also incorporate respiratory information in the form of respiratory modulations and thus, RRV metrics can be derived from heart related signals, which eliminates the need for an additional respiratory signal.

Past research on sleep staging with other physiological signals have utilized publicly available datasets such as Montreal Archive of Sleep Studies (MASS), Sleep Apnea (SA), Sleep Heart Health Study (SHHS), Multi-Ethnic Study of Atherosclerosis Study (MESA) and Computing in Cardiology (CinC) datasets, to validate their proposed methods. Sun et al. [24] proposed a hierarchical sequential network using EOG and ECG signals. They used both hand-crafted features as well as the network trained features for 5-class sleep stage classification and they were able to obtain F1 scores of 0.781 and 0.740 on MASS and SA datasets respectively. Wei et al. [25] used LSTM network based on single lead ECG signal and were able to achieve accuracy of 71.16% and Cohen's kappa statistic of 0.52 on 5-class sleep staging. Meanwhile, Sridhar et al. [26] have developed CNN based architecture and obtained an overall accuracy of 0.77

and Cohen's kappa of 0.66 on a held-out portion of the SHHS dataset.

2.4 Existing Hardware

By analysing the previous studies, we were able to obtain a clear idea on the applications and the limitations that may arise during the hardware development. Following are some of the key factors that are derived from the initial literature study, which are relevant to our project .

- When comparing ear-EEG with scalp EEG, though the raw signal power is significantly less in ear-EEG, the signal to noise ratio (SNR) of both the EEG signals are approximately same. By trading-off the spatial information, a user-friendly, easy to use, and less complex EEG monitoring can be obtained using ear-EEG sensors.
- To this point of time, many studies were conducted and are being conducted in developing hardware designs that would aid to measure ear-EEG signals with least discomfort to the user. The primary focus of the hardware design is to maintaining good electrode-skin contact, self-administration, and least obstruction to the sleep. Furthermore, the EEG signals acquired from the hardware setup should be able to distinguish the sleep stages accurately.
- When observing the existing hardware designs of ear-EEG measurement, they can be classified into 2 main categories namely, earpieces that utilise

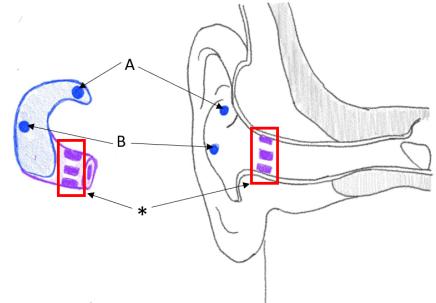


Fig. 2.1: The chosen electrode placements

Wet electrodes. Here, the electrode gel or saline is used between the electrode and the skin to increase the electrode-skin contact. Therefore, the signal strength acquired will be more. So, the signals acquired will be less prone to noise and motion artifacts. But wet electrodes disturb the quality of sleep which will greatly affect the final outcome which is sleep stage classification [27, 28].

Dry electrodes. Earpieces that utilize dry electrodes are more user-friendly and are less likely to hinder the quality of sleep. But it does not have a robust electrode to skin contact. Therefore, these are more susceptible to noise and motion artifacts reducing the overall SNR [29, 30].

- The earpieces can be classified into two other categories, in-ear [27, 28, 30, 31] and outer-ear [29, 32, 33] earpieces. This classification is based on whether the

electrodes are placed inside or outside the ear lobe. Though the signal quality is better in in-ear earpieces, the difference in quality is insignificant.

- PPG signals can also be measured from the ear lobe [34]. Ear-PPG signals can be used to measure the heart rate. Ear-PPG sensors are accurate, non-invasive, and time-efficient [35].
- Apart from ear-EEG and ear-PPG signals, incorporating EOG signals is essential as eye movement is a key indicator for REM and NREM sleep stage classification.
- The shape of the ear differs from person to person anatomically. Therefore to maintain stable contact between the ear and the ear piece, various techniques were used in previous studies such as silicon caps of different sizes, user specific (3D printed or ear imprint) designs [28], size adjustable devices, and memory foams [27].

Chapter 3

SLEEP STAGE CLASSIFICATION ALGORITHM

This section discusses the approaches taken to develop a deep learning based automatic sleep stage classification algorithm, objectives, alternatives and possible future plans. The overview of our algorithm is illustrated in Fig. 3.1.

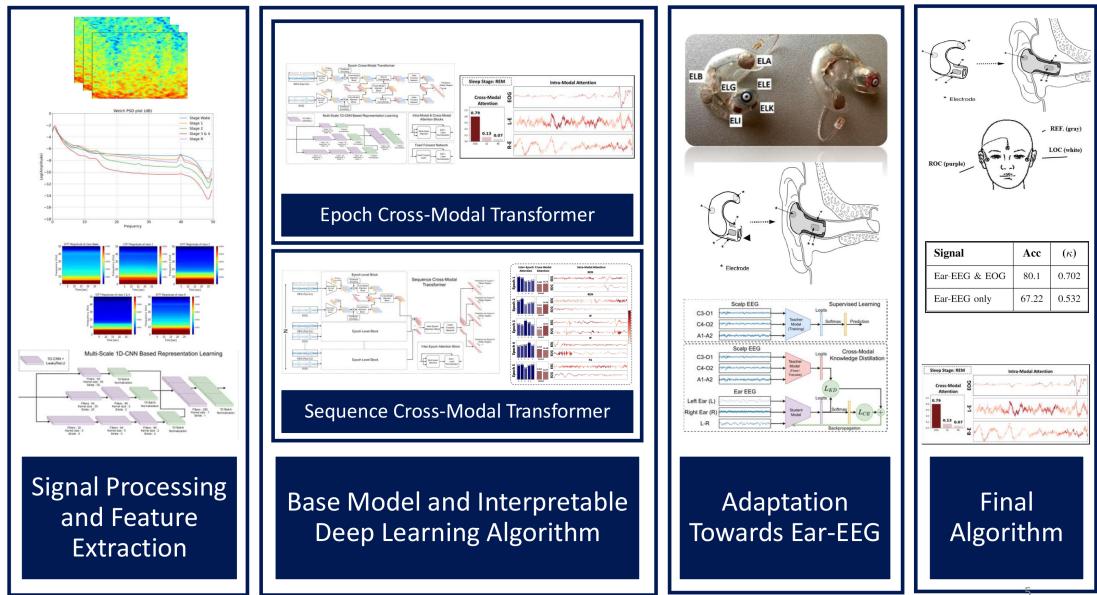


Fig. 3.1: Overview of the sleep stage classification algorithm.

3.1 Signal Processing and Feature Extraction

This section focuses on derivation of handcrafted features from EEG as well as time-frequency analysis of EEG signals. For training our model, initially handcrafted short time Fourier transform (STFT) features were considered. Furthermore, to carry out the feature importance analysis on handcrafted features, a base model of decision tree classifier was developed comprising both STFT and continuous wavelet transform (CWT) based features mentioned under Table 3.1.

In addition, handcrafted features have been used to analyze the EEG data, which can be categorized into three - time domain analysis, frequency domain analysis and time-frequency analysis. Under EEG analysis, fast Fourier transform, periodogram power spectral density (PSD) and Welch PSD were used for frequency domain analysis whereas STFT was used for the time-frequency analysis. As demonstrated in Figure 3.2 and Figure 3.3, all these analyses were able to indicate clear variations

TABLE 3.1: Handcrafted features extracted for feature importance analysis

STFT based Features	Power mean in $\alpha, \beta, \delta, \theta$ bands Power variance in $\alpha, \beta, \delta, \theta$ bands Power skewness in $\alpha, \beta, \delta, \theta$ bands Power kurtosis in $\alpha, \beta, \delta, \theta$ bands Total power in $\alpha, \beta, \delta, \theta$ bands Relative power in $\alpha, \beta, \delta, \theta$ bands Power Ratios : $\delta/\alpha, \delta/\beta, \delta/\theta, \theta/\alpha, \theta\beta/, \alpha/\beta, (\delta + \theta)/(\alpha + \beta)$ bands
CWT based Features	Wavelet energy in $\alpha, \beta, \delta, \theta$ bands Shannon entropy in $\alpha, \beta, \delta, \theta$ bands

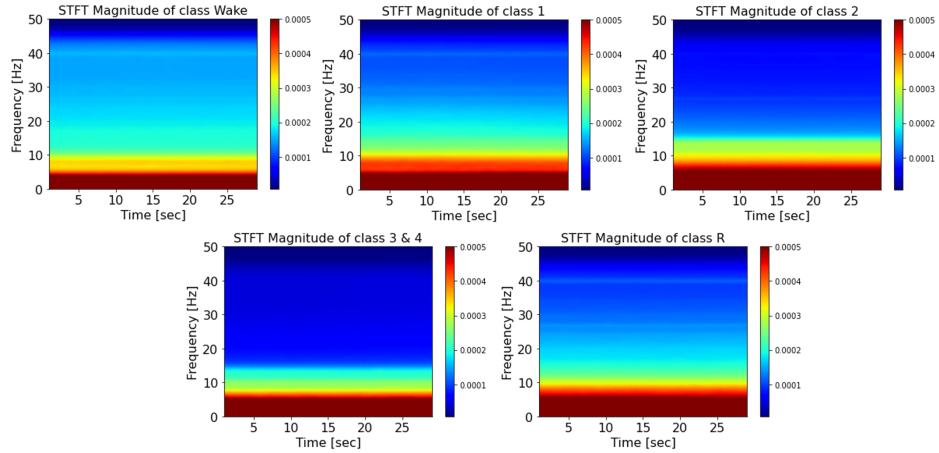


Fig. 3.2: Short time Fourier transform analyses between different sleep stages on the sleep-EDF-expanded dataset.

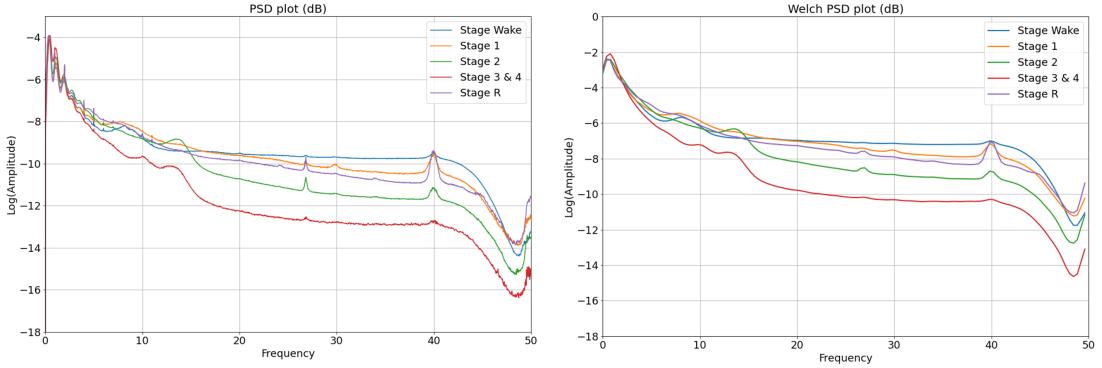


Fig. 3.3: Power spectral density analyses between different sleep stages on the sleep-EDF-expanded dataset.

based on the sleep stages. However, model generated one-dimensional convolutional neural network (1D CNN) features were able to significantly outperform the extracted handcrafted features in both accuracy and kappa values and thus, 1D CNN feature generation was used in our final implementation.

3.2 Base Model and Interpretable Deep Learning Algorithm Development for Sleep Stage Classification

Initially our focus was on building a novel base deep learning model for sleep stage classification using raw EEG and EOG signals from publicly available datasets. The reason is that to compare and validate the performance of our method with the current state-of-the-art methods. We developed a novel deep learning model called, *Cross-Modal Transformers*, for automatic sleep stage classification and it was validated on sleep-EDF-expanded dataset (Sleep-EDF-78) [36] from Physionet [37]. In the following sections, the approaches taken to develop cross-modal transformer are discussed.

3.2.1 Analysis of Alternatives

This section focuses on how we came to the transformers-based solution for automatic sleep stage classification. The sleep stages are sequentially connected and cyclically alternated during sleep, which shows the importance of learning sequential information. Also through literature review and knowledge sharing sessions with sleep experts, we identified that the sleep experts and doctors tends to become unsatisfied due to the black box behavior of the deep learning classification models. They require the validation why an epoch is classified into a certain sleep stage, thus shows the importance of interpretability of the model. We considered several deep learning architectures which could be suitable for sleep staging and analysed them based on following factors:

- Ability to learn sequential information.
- Interpretability.

Convolutional neural networks are the widely using architecture in this domain, and its performance have been saturated in terms of sleep staging [7]. The disadvantages of CNNs are that they are not suitable to learn sequential information and their

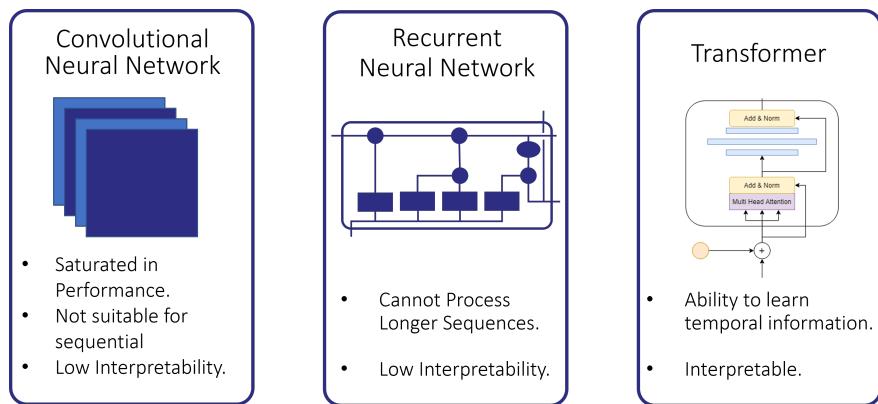


Fig. 3.4: Comparison between deep learning architectures.

interpretability is low. Recurrent neural networks [11, 12] are more suitable to learn sequential information, but they are incapable of learning from long sequences and their interpretability is low. Due to the aforementioned reasons, transformer based deep learning models became more suitable for our task. They are capable of learning representations by utilizing the sequential information in the signals, and they can learn the relationships between a time sample with all the other time samples in the signal. Transformers achieved interpretability because of the multi-head attention blocks in its architecture. We employed transformer based deep learning models to develop automatic sleep stage classification algorithm, because of its advantages.

3.2.2 Epoch Transformer

Initially, we developed epoch transformers, which was a sub section of the SleepTransformer proposed by Phan et. al [15] in their sleep transformer research, which is the only transformer-based sleep classification method in this domain. Epoch transformer as shown in Fig. 3.5, consists of a transformer encoder block and a multi layer perceptron (MLP) block for classification. The architecture of the transformer encoder is similar to the seminal work by Vaswani et al., 2017 [38], where it focuses on learning a

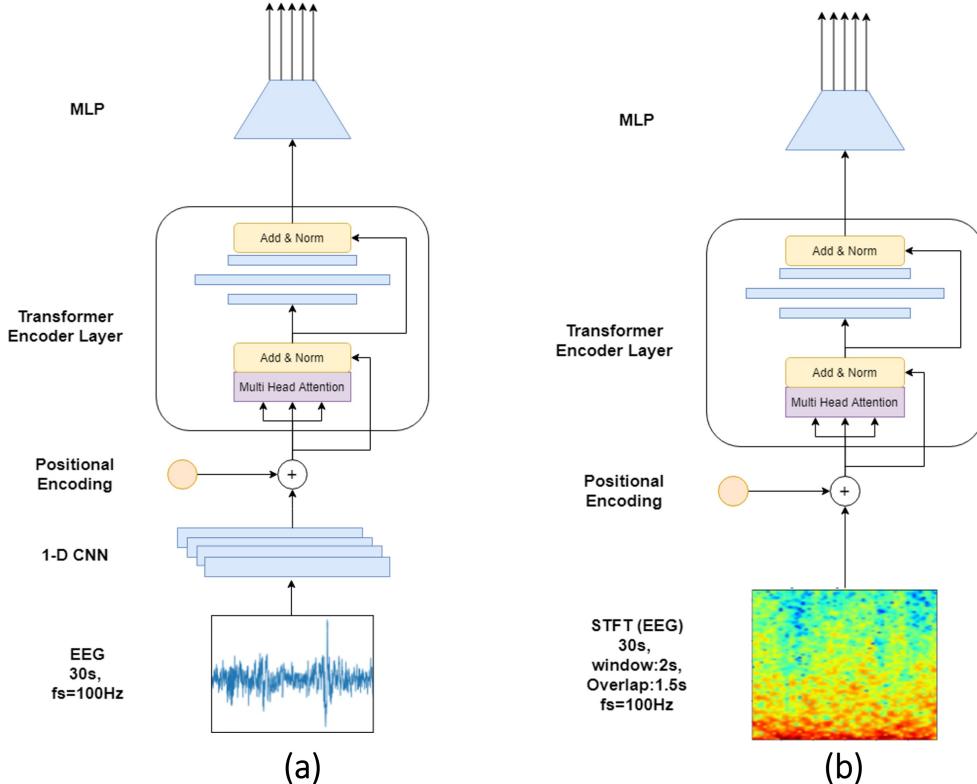


Fig. 3.5: Architecture of the epoch transformer experimented for sleep stage classification on sleep-EDF-expanded 2018 dataset.

representation using the intra-epoch temporal information of a given EEG signal. Then the extracted features are fed into the MLP block for classification.

We experimented the epoch transformer with hand crafted STFT features of an epoch and model-generated 1D-CNN features from the raw EEG signal. Cohen’s Kappa coefficient and accuracy are the two of the evaluation metrics considered for evaluation, and the results are shown in Table 3.2. Through the experiments, we found that the epoch transformer performed better with model generated 1D-CNN features than STFT features. The reason is that 1D-CNNs[39] are capable of extracting optimized features from the raw signal and their parameters such as kernel size, and stride can be adjusted such that there won’t be any loss in temporal information. Therefore, in our final method, we gave importance to 1-D CNN model-generated features.

TABLE 3.2: Performance of the epoch transformer on sleep-EDF-expanded 2018 dataset

Method	Accuracy	Cohen’s Kappa (κ)
Epoch Transformer with 1-D CNN	74.68%	0.6491
Epoch Transformer with STFT	70.82%	0.5938

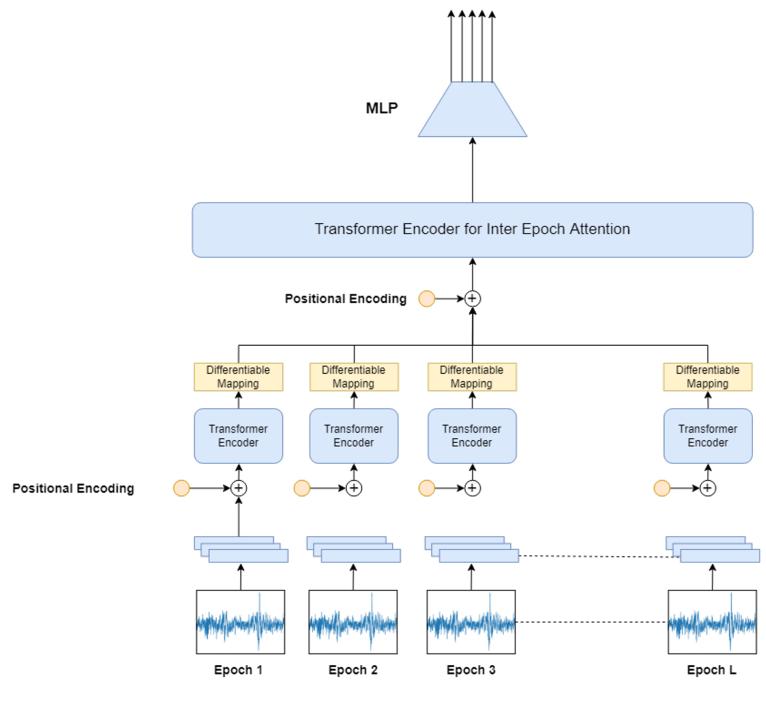
3.2.3 Sequence Transformer

In the manual annotation procedure before annotating an epoch, the sleep expert considers a certain number of epochs before and after the epoch based on AASM guidelines. This indicates the importance of learning the relationship between the epochs for better sleep stage classification. In order to learn these relationships, we implemented sequence transformer shown in Fig. 3.6, which in the first stage that consists of multiple epoch transformers to learn intra-epoch temporal information and an additional transformer encoder block to learn inter-epoch temporal information.

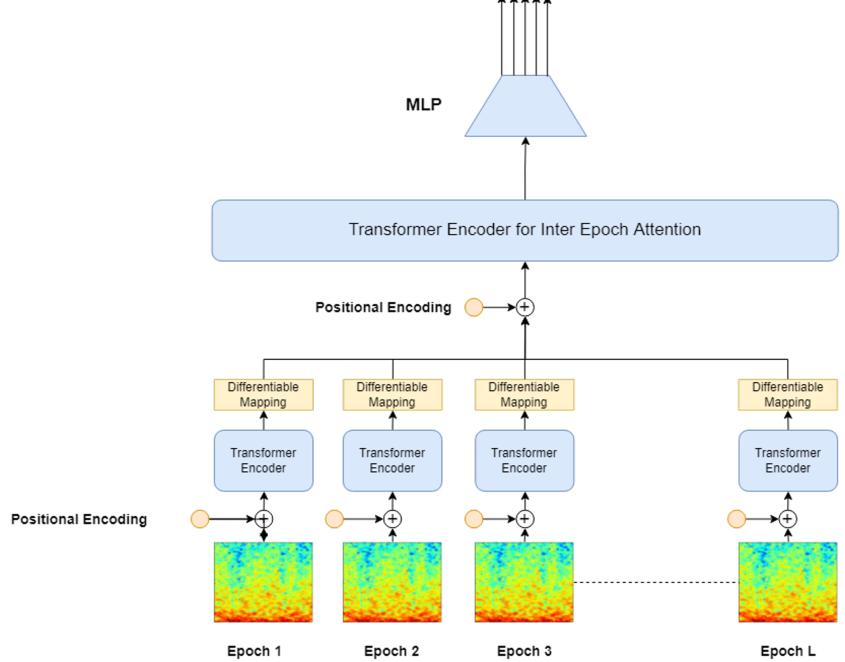
We experimented the sequence transformer with hand crafted STFT features of a epoch and model-generated 1D-CNN features from the raw EEG signal, and evaluated its performance using Cohen’s Kappa coefficient and accuracy. The results given in Table 3.3, shows significant improvement from epoch transformer. The fallback of this method is that it is not capable of learning relationships between different modalities, which makes it challenging to expand for multiple physiological signals.

TABLE 3.3: Performance of the sequence transformer on sleep-EDF-expanded 2018 dataset

Method	Accuracy	Cohen’s Kappa (κ)
Epoch Transformer with 1-D CNN	74.68%	0.6491
Sequence Transformer with 1-D CNN	77.75%	0.6860
Sequence Transformer with STFT	72.99%	0.6289



(a)



(b)

Fig. 3.6: Architecture of the sequence transformer experimented for sleep stage classification on sleep-EDF-expanded 2018 dataset.

3.2.4 Cross-Modal Transformers

This section focuses on our proposed *cross-modal transformers* consisting of a novel cross-modal transformer encoder architecture along with a multi-scale one-dimensional convolutional neural network (1D CNN) for automatic representation learning. The major contributions of the work summarized as follows:

- We propose two cross-modal transformers: 1) Epoch cross-modal transformer and 2) Sequence cross-modal transformer to solve the problem of sleep stage classification under two classification schemes which are one-to-one and many-to-many classification.
- Multi-scale 1D-CNNs are leveraged in our method to learn the feature representations from the raw signals of each modality. This enables our model to learn optimal feature representation to achieve better performance compared to hand-crafted features.
- A novel cross-modal transformer encoder architecture is employed to learn both intra-modal temporal attention, i.e., attention between time steps within a feature representation of a modality, and cross-modal attention, i.e., attention between each modalities. The sequence cross-modal transformer consists of an additional block to learn inter-epoch attention, i.e., attention between adjacent epochs.
- We explain a simple yet effective method which is based on attention mechanisms to enable us to interpret the model as well as prediction. Our proposed method is capable of learning and interpreting: 1) intra-modal relationships, 2) cross-modal relationships and 3) inter-epoch relationships.
- The proposed method was able to achieve state-of-the-art performance in G-mean, sensitivity and specificity. In terms of accuracy, Cohen’s Kappa coefficient and macro F1-score our method was able to achieve performance on-par with the current state-of-the-art.
- Compared to previous methods our method takes less time to train and have smaller model print in terms of parameters.

In the following sections, we formulate the problem definition for one-to-one and many-to-many sleep stage classification task. Then we introduce our cross-modal transformers : epoch cross-modal transformer (shown in Fig. 3.8) for one-to-one classification and sequence cross-modal transformer (shown in Fig. 3.9) for many-to-many classification. We further explain the multi-scale 1D CNN based representation learning and cross-modal transformer encoder under epoch cross-modal transformer.

3.2.4.1 Problem definition

We address the problem of classifying the sleep stage of a 30 s epoch of PSG signals. Our training set with the size of N , consists of labeled 30 s PSG epochs $\{x_i, y_i\}_{n=1}^N$, where $\{x_i, y_i\} \in X \times Y$. Here $X \in R^{T \times C}$ denotes the input space of recorded PSG signals, where T represents the time steps in an epoch and $C \in \{\text{EEG (Fpz-Cz), EOG}\}$ represents the modalities in the recorded PSG signals. $Y \in \{\text{WAKE, N1, N2, N3, REM}\}$ represents the output space of sleep stages. Our goal is to learn a function $f_\theta : X \rightarrow Y$ by minimizing the error $E = E_{(x,y)}(\mathcal{L}(f_\theta(x), y))$ on the given training dataset. Here \mathcal{L} denotes the loss function.

We solve the aforementioned problem using cross-modal transformers under two different classification schemes, which are one-to-one and many-to-many. In one-to-one classification we consider one PSG epoch at a time to predict the corresponding sleep stage [11]. In many-to-many classification, we consider a sequence of PSG epochs at a time and predict their corresponding sleep stages at once [11].

3.2.4.2 Dataset

We used the publicly available sleep-EDF-expanded dataset (Sleep-EDF-78) [36] from Physionet [40], to evaluate our architecture for sleep stage classification. We used the sleep cassette (SC) dataset of sleep-EDF 2018, which consists of 153 whole night PSG recordings from 78 healthy individuals. Each record comprises of two EEG signals (Fpz-Cz and Pz-Oz), a horizontal EOG signal and a submental chin EMG signal. The EMG signal was sampled at 1 Hz whereas all the other signals were sampled at 100 Hz. The PSG recordings were accompanied with hypnograms annotated by sleep technician based on Rechtschaffen and Kales (R&K) guidelines [41]. In the hypnogram, each 30 s epoch of the recorded signal is assigned to one of the following labels : Stage 0, Stage 1, Stage 2, Stage 3, Stage 4, REM, movement time and ‘?’ (unscored). For this study, we converted these annotations to the AASM standards [42], by combining Stage 3 and Stage 4 to a single stage N3 while Stage 0, Stage 1 and Stage 2 are relabelled as Wake, N1 and N2, respectively. Also, epochs consisting of annotations ‘movement time’ and ‘?’ were discarded. Altogether 415,465 30 s-epochs of PSG recordings were extracted.

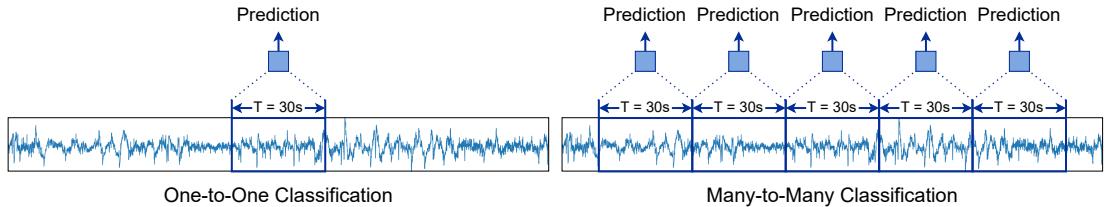


Fig. 3.7: One-to-one and many-to-many classification schemes.

3.2.4.3 Data Preprocessing

Among the extracted epochs, 69% comprises of wake stage. This is due to the existence of long wake stages at the beginning and end of each dataset. Thus, to reduce the redundancy of wake stage, we only considered 30 minutes of wake stages before and after the sleep periods. In the resultant dataset, number of epochs is dropped to 196,350. For our experiments, we have considered both Fpz-Cz EEG signal and horizontal EOG signal. All these signals are initially segmented to a sequence of 30s epochs and corresponding labels are assigned based on the provided annotations. The EEG signal is bandpass filtered between 0.2 Hz to 40 Hz using a zero-phase finite impulse response filter with Hamming window. Then, the signals are normalized such that each signal has zero mean and unity variance.

3.2.4.4 Epoch cross-modal transformer

In this section, we focus on solving sleep stage classification with one-to-one scheme using epoch cross-modal transformer. The proposed epoch cross-modal transformer as shown in Fig. 3.8, consists of two main blocks: 1) multi-scale 1D-CNN for automatic representation learning and 2) a novel cross-modal transformer encoder architecture to learn both intra-modal temporal attention and cross-modal attention. The epoch transformer initially learns two separate feature representation for the input EEG and EOG signals in an epoch using 1D-CNN. Here, features are learned and extracted using non-overlapping windows in the signals. Then, the cross-modal transformer encoder learns a representation by considering intra-modal temporal attention and cross-modal attention, which is then fed into a linear layer for classification. The two main blocks of the epoch transformer are further elaborated in proceeding sections.

3.2.4.5 Multi-scale 1D-CNN for representation learning

Inspired by vision transformers using image patches as sequential data[43], we employ multi-scale 1D-CNN as shown in Fig. 3.8 to learn the feature representation of non-overlapping windows with the size of 0.5s in the raw input signals. Here, the raw 1D signals $X_c \in R^T$ will be mapped into a feature space of $X'_c \in R^{(T/(0.5 \times f_s)) \times E}$, where $c \in C$, f_s is the sampling frequency and E is the embedding size. Let $T/(0.5 \times f_s)$ be T' for simplicity. The features are extracted from non-overlapping windows instead of overlapping windows to improve interpretability, and such that the extracted feature vectors can be fed into the cross-modal encoder as sequential data to learn the attentions between all windows.

We hypothesize that the global and local features in a window of raw signals will contribute towards the classification of the sleep stages. In order to extract both global

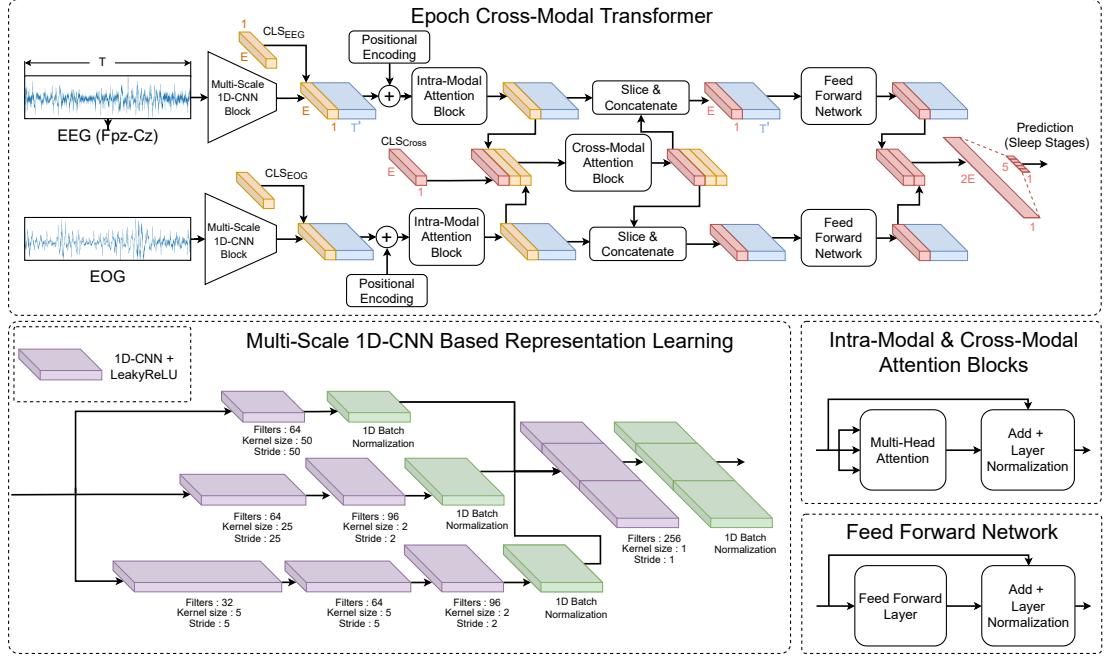


Fig. 3.8: Architecture of proposed epoch cross-modal transformer

and local features in a window, we employed multi-scale 1D-CNN, where the raw input signal will go through three parallel paths as shown in Fig. 3.8: 1) 1D-CNN with kernel size of 50, 2) 2 1D-CNNs with kernel sizes of 25 and 2 respectively, and 3) 3 1D-CNNs with kernel sizes of 5, 5 and 2. Each 1D convolution layer will be followed by a LeakyReLU activation. The features extracted at different scales will be normalized using batch normalization. Finally, the extracted feature representations will be concatenated along the embedding dimension and will undergo an additional 1D-CNN with kernel size of 1, followed by LeakyReLU activation and batch normalization.

3.2.4.6 Cross-modal transformer encoder and classification

We propose a novel cross-modal transformer encoder architecture to learn a powerful feature representation by attending intra-modal temporal information and cross-modal relationships. As illustrated in Fig. 3.8, the cross-modal transformer encoder consists of two main blocks: 1) attention and 2) position wise fully-connected feed-forward networks blocks.

Initially, for each modality c , a learnable CLS_c vector $\in R^{(1 \times E)}$ similar to the one proposed in BERT [44] is randomly initiated and concatenated with the output of the multi-scale 1D-CNN block along the time axis. Similar to the seminal work [38], positional encodings are added to the concatenated vector and fed to the intra-modal attention block, to learn the relationships between all time steps in the modality. We only extract the vector representation corresponding to the CLS_c vector from the output from each modality, as it aggregates all the information of the intra-modal tem-

poral information [44]. A new learnable CLS vector named CLS_{cross} is randomly initialized and concatenated along with the CLS_c vector representation extracted from each modality. Cross-modal attention block is employed on the aforementioned vector to learn the relationships between the modalities, which get aggregated in the corresponding representation of CLS_{cross} . The vector representation corresponding to the CLS_{cross} is extracted and CLS_c representation of each modality is replaced by this vector. Then the concatenated vectors are passed through the feed-forward network. Finally, the vector representations corresponding to CLS_{cross} in each modality are extracted, flattened and pass through a single linear classifier with 5 neurons for classification. The operations of the intra-modal attention block, cross-modal attention block and feed-forward network are similar to the seminal work by [38]. The attention blocks consist of multi-head attention followed by residual addition and layer normalization. The feed-forward network consists of a feed-forward layer followed by residual addition and layer normalization. An additional linear layer is sufficient for classification of the sleep stages from the extracted representation, which shows the power of the learned representation.

3.2.4.7 Sequence cross-modal transformer

In order to solve sleep stage classification with many-to-many scheme we employ sequence cross-modal transformer, which is an extension of the previously explained epoch cross-modal transformer as shown in Fig. 3.9 . Here, the proposed architecture consists of multiple epoch level block for each 30s epoch of raw PSG signals in a sequence. The epoch level block is build using multi-scale 1D-CNN block and intra-modal attention block for each modality and a cross-modal attention block. The cor-

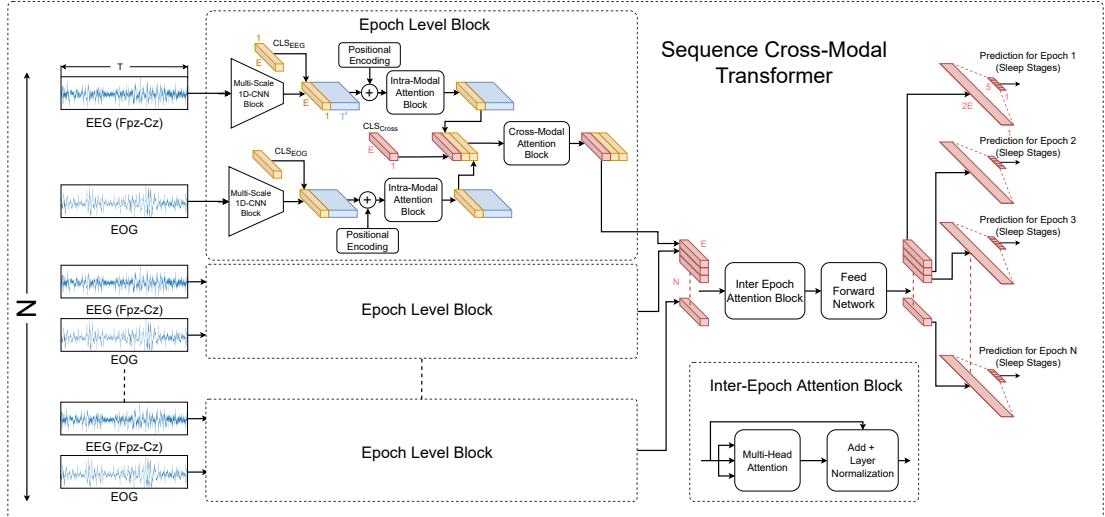


Fig. 3.9: Architecture of proposed sequence cross-modal transformer

responding vector representation of CLS_{cross} vector of each epoch block is extracted and concatenated. Inter-epoch attention block followed by a feed-forward network is employed on the concatenated vector to learn the relationships between the epochs. Finally, we extract and flatten output representation corresponding to each epoch separately and employ a linear classifier to predict their sleep stages. During inference, we let the model run across the sequence of PSG epochs, where multiple predictions for an epoch will be achieved. Then, we calculate the mean probability to predict the correct sleep stage.

3.2.4.8 Interpretability

In this section, we propose a simple method to interpret the predictions of cross-modal transformer, which would be beneficial for the sleep experts and would increase their confidence in the system. The main advantage of our architecture is that the attention mechanism can be simply utilized to learn and interpret: 1) intra-modal relationships, 2) cross-modal relationships and 3) inter-epoch relationships.

We extract the output of the intra-modal attention block and calculate scaled dot product between the representation corresponding to CLS_c and representations corresponding to the non-overlapping windows of the raw signal to learn intra-modal relationships. This gives attention weights for each non-overlapping window of the raw signal, which interprets their impact on the prediction. The relationships between modalities is interpreted by calculating the scaled dot product between the representations in the output of cross-modal attention block corresponding to CLS_c of each modality. Finally, we interpret the inter-epoch relationships by getting the scaled dot products between CLS_{cross} of each epoch.

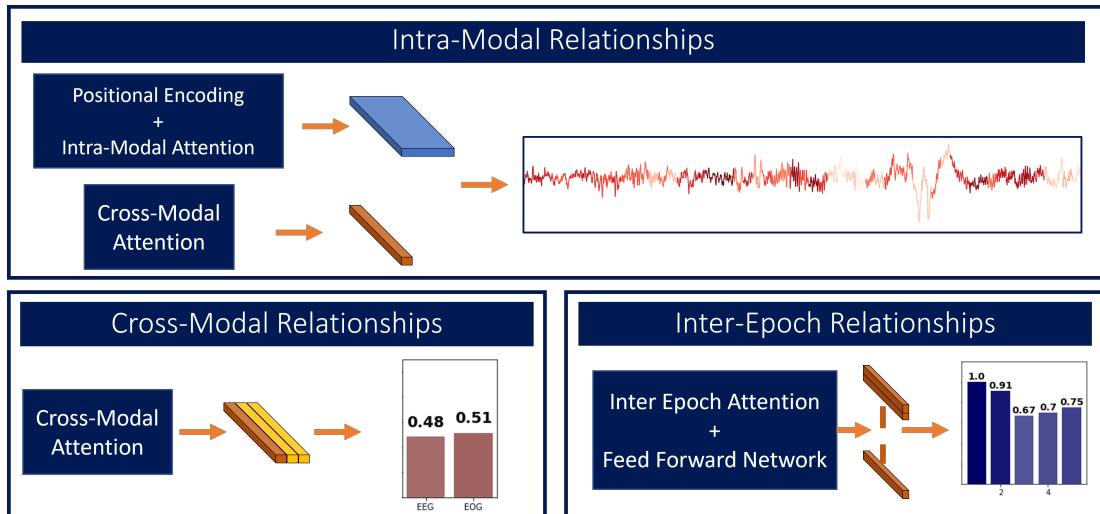


Fig. 3.10: Method for interpretation

3.3 Adaptation of Algorithm Towards Ear-EEG

PSG signals are the a golden standard for sleep staging. But PSG signals based sleep study can induce discomfort to the subject, cannot be used in home settings, as they are usually a complex setup. Ear EEG is a potential candidate for sleep stage detection in home settings, as its relatively simple setup and more comfortable than scalp EEG, as scalp EEG usually involves scalp EEG cap with multiple electrodes.

In ear-EEG domain, as discussed in literature, there are few early works, which had shown that Ear EEG is a potential candidate for sleep staging. Also few other works had proven the correspondence between scalp EEG and ear EEG signals. In ear EEG base sleep staging research domain, there are plenty of works focusing on comfortable ear EEG acquisition setups such as designing different ear molds, flexible ear-EEG electrodes setups. But there are few or no works focusing on algorithms aspects of the Ear EEG. Our study will explore the hardware aspect of ear-EEG based sleep staging.

3.3.1 Dataset

In our study, we utilized a dataset from [45], consisting one whole night sleep recordings from 9 healthy subjects (age : 26-44; 3 female and 6 male). The dataset contains simultaneously recorded PSG signals and ear-EEG signals. The PSG signals consists of 8 scalp recordings - O1, O2, C3, C4, A1, A2, F3 and F4 according to the international 10-20 system, 2 EOG recordings - LOC and ROC and a chin EMG recordings. Ear EEG consists of 12 ear electrode recordings, where 6 recordings from each ear - ELA, ELB, ELE, ELI, ELG, ELK, ERA, ERB, ERE, ERI, ERG and ERK. The sampling rate of the data is 200Hz. The dataset was manually scored based on the international AASM guideline. For further details of the dataset, we direct the readers to [45]. In the dataset, a single epoch which labeled as ‘Unscored’ was discarded.

3.3.2 Data preparation

All the recordings (scalp and ear EEG) were $0.2 - 42\text{Hz}$ band-pass filtered. The EOG and EMG channels are not considered in our study. The scalp EEG and ear-EEG derivations are extracted from the recording as follows:

Scalp EEG : Three scalp-EEG derivations are extracted from 8 scalp-EEG channels : $C3 - O1$, $C4 - O2$ and $A1 - A2$.

Ear EEG : Three ear-EEG derivations are extracted from the 12 ear-EEG channels. [45]. (\bar{x}) denotes mean.

$$L_1 = \overline{ELA, ELB, ELE, ELI, ELG, ELK}$$

$$R_1 = \overline{ERA, ERB, ERE, ERI, ERG, ERK}$$

$$L - R = L_1 - R_1$$

$$L = \overline{ELA, ELB} - \overline{ELE, ELI, ELG, ELK}$$

$$R = \overline{ERA, ERB} - \overline{ERE, ERI, ERG, ERK}$$

Here $L - R$ derivation gives the electric potential difference between left and right ears. L and R derivations give the potential difference between concha and ear canal in left and right ear respectively. The correspondence between scalp EEG derivations and ear EEG derivations can be observed from Figure 3.11.

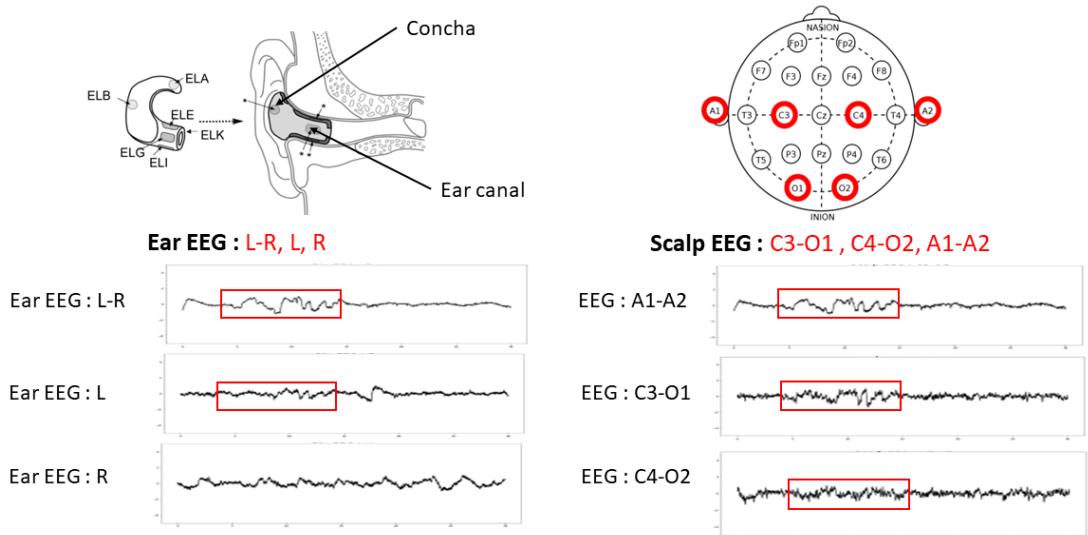


Fig. 3.11: The correspondence between scalp and ear EEG derivations. Left - ear EEG derivations (L-R, L, R) and Right - scalp ear EEG derivations (A1-A2, C3-O1, C4-O2)

Channel rejection was done as some of the ear-EEG recordings were noisy. All intra-ear derivations are calculated, and the mean power (P_{ij}) within the frequency range of 10-35Hz was calculated [45]. let $m_i = \text{median}(P_{ij}, \forall j)$. An electrode i was rejected if m_i is an outlier. Altogether 15 channels were rejected from the dataset. Recordings of subject 5 was removed from the dataset as both the ear canal channels (ERA and ERB) from right ear are rejected, thus mean cannot be calculated. In Mikkelsen et al., missing channels were substituted with a copy of another subjects recordings [45].

TABLE 3.4: Description of the rejected channels

Subject No	#1	#2	#3	#4	#5	#6	#7	#8	#9
Rejected electrodes	2	2	1	2	5	0	0	1	2
Usable electrodes	10	10	11	10	7	12	12	11	10
Scored epochs	1039	964	932	1150	-	926	758	293	857

3.3.3 Supervised Learning on Ear-EEG

The novel cross-modal transformer was initially designed for the scalp-EEG and EOG signals from sleep-EDF-expanded dataset. The model was able to yield competitive metrics with the state-of-the-art methods.

Following the development of novel cross-modal transformer architecture the focus was given to adapt the developed models towards the ear-EEG dataset. Initially, the model was trained with the 3 ear-EEG derivations (L-R, L-E, R-E) and manual sleep annotations as input in a supervised manner using the novel epoch cross-modal transformer. The results of the supervised training reported in Table 3.6. The epoch cross-modal transformer was able to achieve average accuracy of 67.28 and kappa of 0.532, which is higher than the result produced by Mikkelsen et al., 2017 [45] which utilizes the random forest classifier on the same dataset.

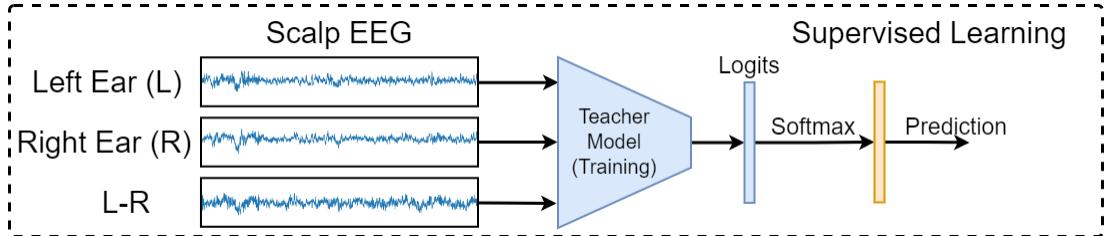


Fig. 3.12: Supervised Training Framework

TABLE 3.5: Performance comparison of Epoch Cross-Modal Transformer on ear-EEG Dataset with state-of-the-art

Subjects	1	2	3	4	6	7	8	9	Avg
Kappa of [45]	0.05	0.36	0.57	0.60	0.59	0.75	0.65	0.44	0.50
Kappa of Ours	0.232	0.427	0.680	0.575	0.632	0.778	0.475	0.458	0.532

The model was trained and tested based on leave one subject out (LOSO) approach, where the model will be trained with the data of 7 subjects and tested on the data of the remaining 1 subject.

3.3.4 Transfer Learning

Transfer learning is a widely used technique in computer vision domain. In transfer learning, we pre-train a model using scalp EEG dataset, and fine tune it on ear-EEG dataset. This approach didn't improve the performance of sleep staging with ear-EEG dataset. This could be due the difference in amplitude and difference in spatial location of signal acquisition method. Although there is correspondence between ear-EEG and scalp-EEG its not appropriate to map the feature map of them, but its much more sensible to map the probability distribution of model trained on ear-EEG and scalp-EEG. The results of Transfer learning is reported in Table 3.6.

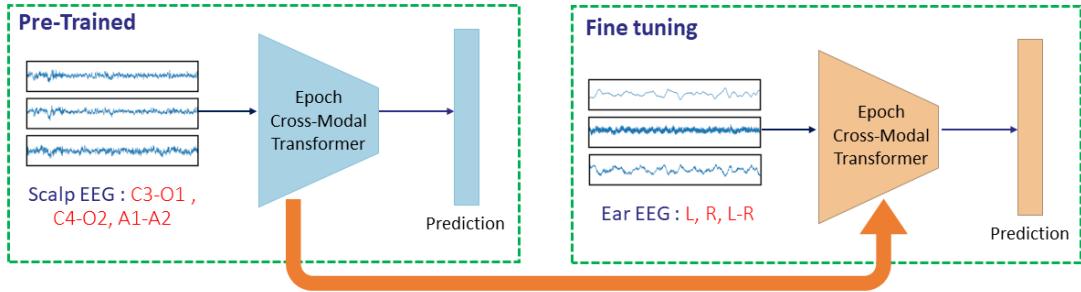


Fig. 3.13: Transfer Learning Framework

3.3.5 Knowledge Distillation

Knowledge distillation [46] is a process of transferring knowledge from a teacher model to a student model. In knowledge distillation the student model will learn to mimic the teacher model by leveraging the knowledge embedded in teacher model to obtain similar or high accuracy. In our study we used cross-modal knowledge distillation, where the teacher model is trained on scalp EEG modality and its knowledge distilled into the student model that learns knowledge from a different modality, i.e., ear-EEG.

In our study, we clone of the base sleep stage classification architecture : teacher and student models. Both teacher and student models will have same base architecture. There are different types of knowledge distillation approaches such as response based knowledge distillation and feature based knowledge distillation. In response based knowledge distillation, the last logit layers of teacher and student models are mapped. whereas in feature based knowledge distillation intermediate feature maps are also mapped between both teacher and student models. During ablation studies, the response based knowledge distillation approach was yielding better results.

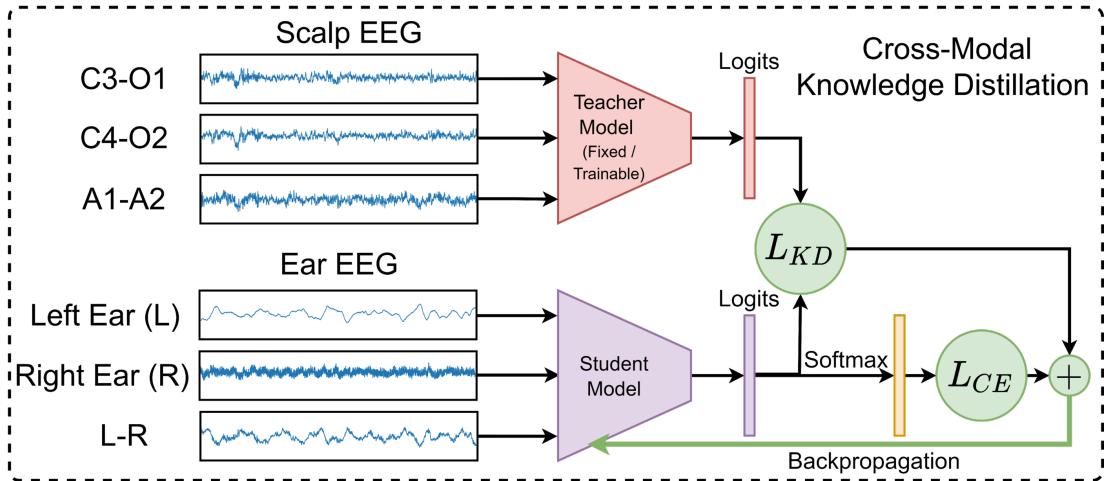


Fig. 3.14: Knowledge distillation Framework

The cross-entropy loss between prediction of teacher model and true targets is used as the loss function of teacher model. The cross entropy loss between prediction of

student model and true targets used along with mean squared error between soft targets of teacher model and predictions of student model as loss function of student model.

In our study we experimented with 2 training strategies : online knowledge distillation and offline knowledge distillation.

3.3.5.1 Online Knowledge Distillation

In online distillation approach both the teacher and student models are trained simultaneously, whereas the teacher model will learn to predict the true label and the student model learns to predict the true targets as well as the distribution of the teacher model. The online knowledge distillation training strategy requires high computational resources compared to offline knowledge distillation training strategy.

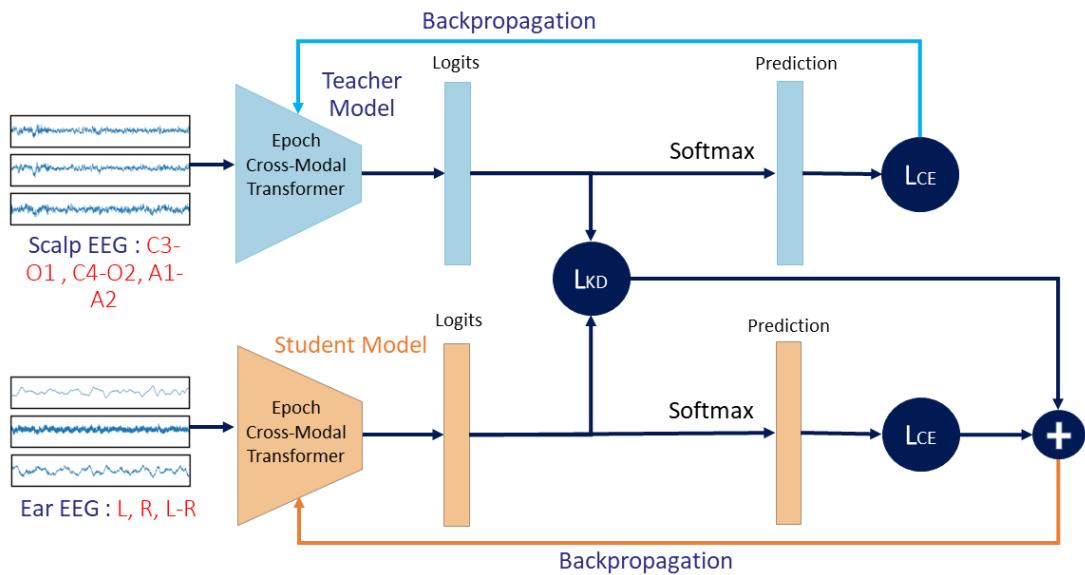


Fig. 3.15: Online Knowledge distillation Framework

3.3.5.2 Offline Knowledge Distillation

In offline distillation approach, the teacher model is pre-trained on scalp EEG dataset. During knowledge distillation training the weights of the teacher model is fixed and used to guide the student model. In this approach the student model will learn the distribution of the teacher model, and thus will mimic the predictions of teacher model.

3.3.6 Training Setup

The model is trained using Adam optimizer with learning rate (lr), β_1 and β_2 set to, 10^{-3} , 0.9 and 0.999, respectively. The batch size is experimentally chosen as 32. We trained the supervised model and knowledge distillation model for 100 epochs.

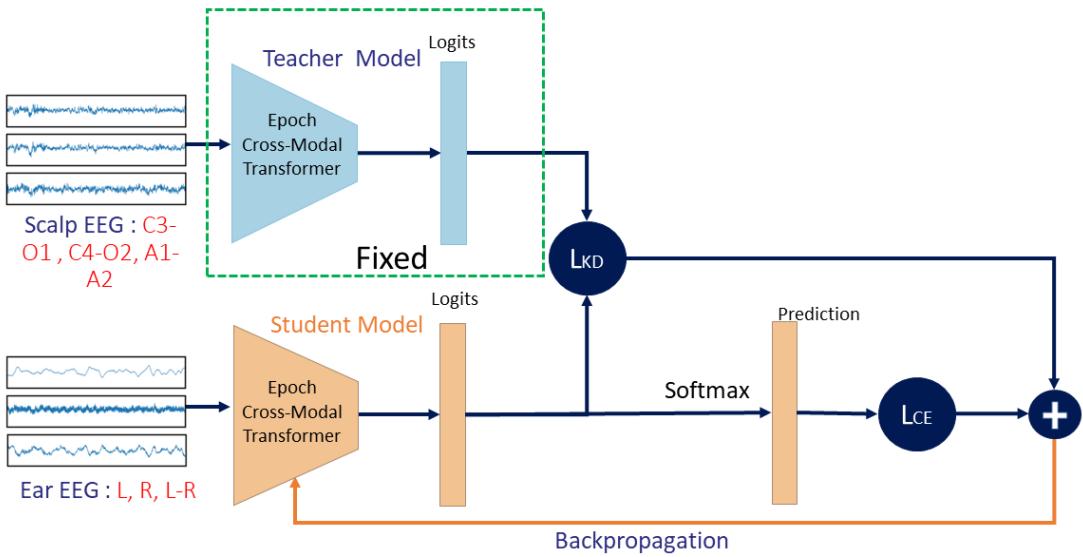


Fig. 3.16: Offline Knowledge distillation frame work

The dataset was partitioned in leave-one-out fashion, as this is a most probable real-application scenario : The model was trained on dataset of 7 subjects and tested on 1 subject, which the model has not seen during training. The model is implemented in Pytorch environment and trained using a Nvidia Quadro RTX 5000 graphics card with 16 GB memory.

3.3.7 Results

We used accuracy (ACC), Cohen’s kappa coefficient (κ) as the performance evaluation metrics of our method. Results of knowledge distillation approach is reported in Table 3.6, the class-wise performance of epoch cross modal transformer on ear EEG dataset is reported in Table 5.2.

TABLE 3.6: Results of Transfer Learning and Knowledge Distillation

Methods	Metrics	Subject 1	Subject 2	Subject 3	Subject 4	Subject 6	Subject 7	Subject 8	Subject 9	Average
Scalp EEG	Accuracy	68.15	68.04	79.48	82.18	81.23	82.60	77.69	85.49	78.11
Ear EEG		41.38	65.83	78.13	73.94	75.80	85.23	58.63	58.87	67.28
TF		31.14 ↓	60.99 ↓	81.35 ↑	72.54 ↓	72.33 ↓	82.66 ↓	63.94 ↑	45.76 ↓	63.84 ↓
KD (Offline)		49.18 ↑	62.60 ↓	78.33 ↑	76.18 ↑	79.16 ↑	86.02 ↑	66.44 ↑	58.97 ↑	69.61 ↑
KD (Online)		46.31 ↑	64.42 ↓	83.23 ↑	77.94 ↑	75.60 ↓	84.91 ↓	63.94 ↑	57.17 ↓	69.19 ↑
Scalp EEG	Kappa	0.539	0.464	0.702	0.711	0.711	0.736	0.699	0.797	0.670
Ear EEG		0.232	0.427	0.680	0.575	0.632	0.778	0.475	0.458	0.532
TF		0.128 ↓	0.348 ↓	0.734 ↑	0.555 ↓	0.580 ↓	0.746 ↓	0.540 ↑	0.220 ↓	0.481 ↓
KD (Offline)		0.259 ↑	0.341 ↓	0.687 ↑	0.620 ↑	0.680 ↑	0.790 ↑	0.571 ↑	0.456 ↓	0.551 ↑
KD (Online)		0.263 ↑	0.403 ↓	0.747 ↑	0.647 ↑	0.617 ↓	0.772 ↓	0.543 ↑	0.417 ↓	0.551 ↑

TABLE 3.7: Performance comparison between supervised training and knowledge distillation

Method	Per-class Performance				
	W	N1	N2	N3	REM
Scalp EEG	83.89	6.26	88.68	70.06	79.41
Ear EEG	64.94	6.46	80.15	70.28	60.37
KD (Offline)	60.57 ↓	2.02 ↓	85.94 ↑	66.30 ↓	66.54 ↑
KD (Online)	68.42 ↑	3.03 ↓	81.75 ↑	75.80 ↑	58.29 ↓

The results are clearly validating that the cross-modal knowledge distillation helps to improve the performance of ear EEG based sleep staging by utilizing the knowledge on scalp EEG. In our study we show that, with cross modal knowledge distillation from scalp EEG domain to ear EEG domain improves the accuracy and kappa of Ear EEG based sleep staging by 3-4%.

3.3.8 *Final algorithm*

The ultimate goal of the project to develop a sleep staging algorithm to classify sleep stages using ear EEG and EOG signals. In alignment with our final objective we adapted the developed epoch cross-modal transformer for ear EEG and EOG signals. L and R ear EEG derivations were used and the EOG was derived from both EOG electrodes (ROC - LOC). We used accuracy (ACC), Cohen's kappa coefficient (κ), macro averaged F1-score (MF1), and macro averaged G-mean (MGm) as the performance evaluation metrics of our model. The results of ear EEG and EOG based sleep staging is reported in Table 3.8.

TABLE 3.8: Performance of Ear-EEG and EOG based sleep staging

Signals	ACC	Kappa	MF1	MGm
Ear EEG and EOG	80.1	0.702	0.654	0.799
Ear EEG only	67.22	0.532	0.519	0.708

The confusion matrix of the ear EEG based sleep stage classification and ear EEG and EOG based sleep stage classification is given in 3.17. When EOG signal is incorporated F1 score of all 5 classes are improving significantly. EOG signal has improved the Wake class prediction by 23%, REM class prediction by 22%, and N1 class prediction by around 12%.

Interpretation of the classification results is indicated in 3.18.

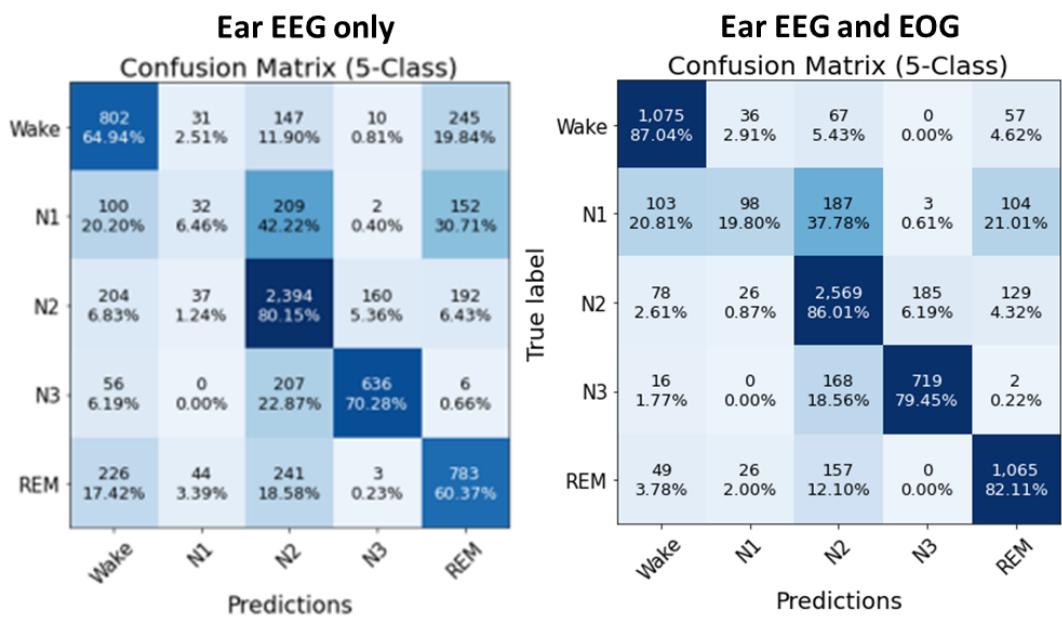


Fig. 3.17: Confusion matrix of ear EEG only (left) and ear EEG and EOG signal (right)

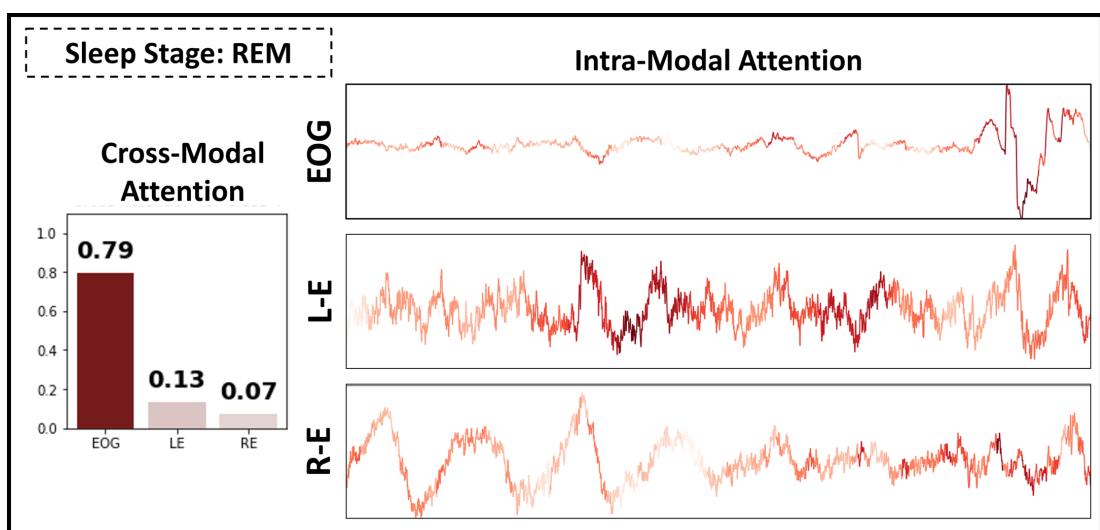


Fig. 3.18: Interpretation with ear EEG and EOG signals

Chapter 4

SIGNAL ACQUISITION HARDWARE SETUP

This section covers the step by step implementation of developing the hardware required for the comfortable and accurate ear-EEG and EOG signal acquisition. Initially, number of modalities and the features that can be implemented into our project scope were finalized, through literature review. From the gained knowledge, a custom made ear mold was developed that is both stable and comfortable to the user. The designed earpiece is capable of acquiring both ear-EEG and EOG signals from the ear parallelly. A 3D enclosure model was designed to facilitate taking all the wires behind the ear and to the signal acquisition device. The signal acquired by the signal acquisition setup is analysed and it correlated with expected signal.

4.1 Initial Study

A thorough literature study was conducted initially so that all the previous studies that have been conducted related to our own study is analysed. From the initial study phase, we finalized some of the key features of the earpiece to be designed. We have planned to use Ag-AgCl silver pellet electrode as they are small in size and can be incorporated in the earpiece without compromising the user comfort level. **We were successful in selecting the appropriate electrode placement positions based on the previous works, which would give us significant spatial information with less noise addition.**

As per the developed algorithm, we would be needing 3 types of ear-EEG signals, namely left polarized signal and right polarized signal which are commonly termed as intra-ear EEG signals and the left to right polarized signal (inter-ear EEG signal). Therefore the electrode numbers and placements are chosen such that both inter-ear and intra-ear signals can be simultaneously obtained. **The electrode placements are, one electrode in the ear canal and one electrode in the Concha area. In order to reduce motion artefacts and noise, we have used in-ear wet electrode earpiece designs.**

4.2 Preliminary Design

For the preliminary design, to test the concept of ear-EEG and the other modalities chosen regarding the development of the signal acquisition system, an ear mold was made using ear impression material (Otoform) which is a custom fit modal to the user's ear. This ear impression gives good electrode to skin contact without any discomfort to

the ear. We collaborated with Wickramarachi Institute of Speech and Hearing (WISH) to obtain these ear impressions.

Although initially we have planned to go ahead with OpenBCI for signal acquisition, we were unable to get the OpenBCI connected to the computer because of the malfunctioning of the OpenBCI (USB) BLE dongle. Therefore we went ahead with the Biosignal amplifier device available in the Biomedical Lab.

4.3 Overview of the Hardware System

This section covers the overview of the hardware aspects of the signal acquisition system, from signal acquisition through the electrodes to validating the acquired signals. The intended plan of the hardware development was not met in the given time due to the global pandemic and economic crisis faced by Sri Lanka. Instead alternative methods were followed to reach the end goal successfully.

4.3.1 The Intended system

This section focuses on the intended hardware signal acquisition system, and Fig. 4.1 illustrates the simplified block diagram of the overall system. As per the plan, the ear-EEG signal is acquired using silver pellet electrodes and carried to the system through wire-harness which can avoid addition of electromagnetic waves and other external interferences with the acquired biosignal through the wire. AD8244 is used for active shielding and then the acquired signal undergoes low pass filtering. Then the resultant signal is fed to ADS1299 analog front end for both amplification and analog to digital conversion. We employ STM32 micro-controller for controlling the overall signal acquisition system.

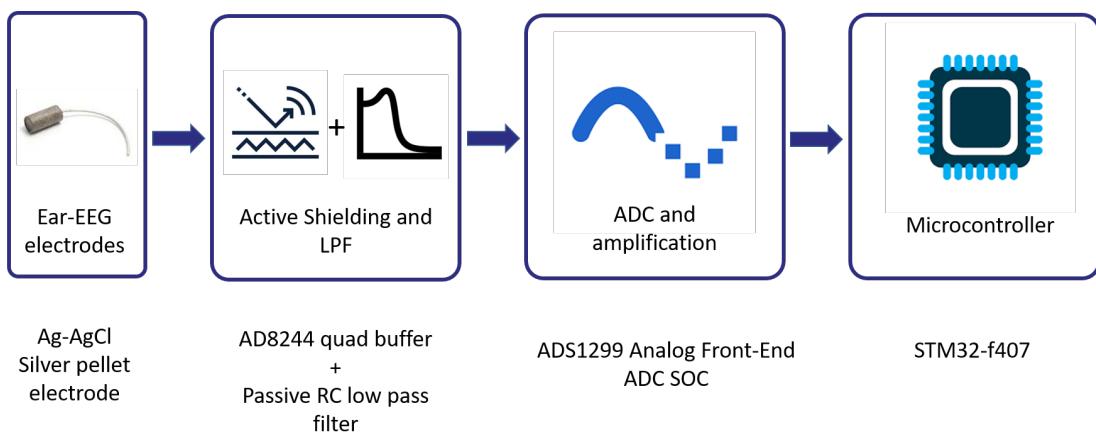


Fig. 4.1: The overall block diagram of the intended ear-EEG acquisition device

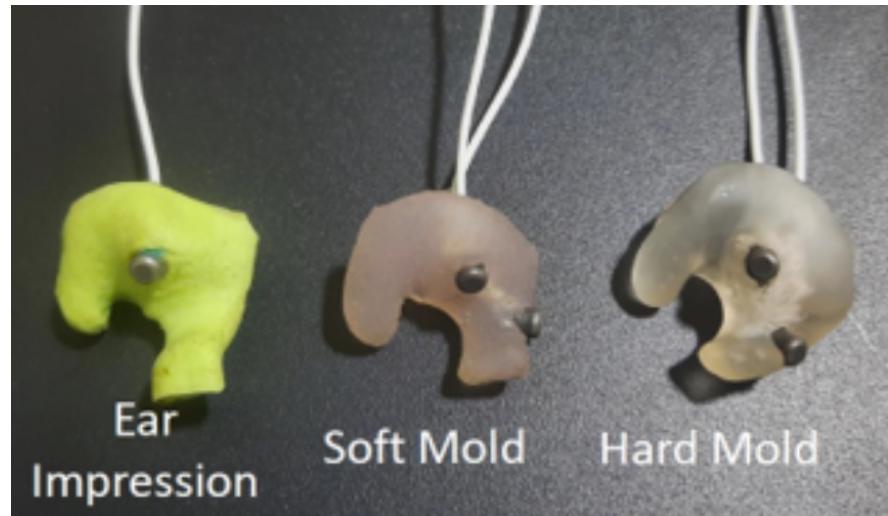


Fig. 4.2: Types of ear molds

4.3.2 *The Earpiece*

By collaborating with the Wickramarachi Institute of Speech and Hearing (WISH), custom made ear impressions (Refer Fig. 4.2) were obtained which is ideal for rigid skin to electrode contact while maintaining optimum comfort to the user. As per the user's wish, a soft mold or hard mold earpieces can be obtained. The choice will not affect the signal quality in any way. The impression material used for these ear molds is Egger. We have placed 2 electrodes in the ear, one in ear canal and one in Concha which will help to get signals with left or right polarity and left to right polarity, which is needed for our algorithm as explained under the algorithm development section.

As we are not using EOG signals, check whether we can improve electrode positions

Furthermore, through the electrodes placed at these particular points, we were able to acquire EOG signals and occasional ECG signals apart from ear-EEG signals. The ability to acquire both EEG and EOG signals from the same electrodes in the mold, considerably reduced the complexity of the system and increased the user comfort. Thus, the need for additional attachment of EOG sensors in the outer temple region of the forehead is avoided. We have not utilized the ECG signals obtained from the same ear mold because obtaining the ECG signals from any ear mold from any patient is uncertain. The occurrence of ECG signals in the acquired biosignal depends on the presence of major artery in the ear canal and placement of the electrode on top of the artery.

Though the orders were made for ear-EEG Ag-AgCl silver pellet electrodes, without completely relying on them, custom electrodes were made by deforming the button electrodes. And with the ear molds obtained from Wickramarachi, we were able to acquire good quality EEG and EOG signals.



Fig. 4.3: The CAD and printed enclosure design

4.3.3 *The enclosure design*

The CAD model in Fig. 4.3 is the enclosure model that is designed to take all the wires from the mold behind the ear and to the acquisition device with least disturbance to the user while sleeping. Apart from the enclosure, a ninja flex (TPU or thermoplastic polyurethane) material is added between the enclosure and the skin to avoid possible skin irritations. Since the ninja flex material is flexible and soft, it will increase the stability of the enclosure while sleeping.

Before coming to this specific design, many other alternative designs were analysed. The design in Fig. 4.4 is another design that was prepared, which was inspired from Muse-2 bands. After collaborating with integrated design department students and analysing the practicality and user comfort, the design in Fig. 4.3 which was inspired from the standard hearing aid designs was developed. This is the least complex, most comfortable, and the most stable design compared to the alternative designs considered.



Fig. 4.4: The enclosure design inspired from Muse 2 bands

4.3.4 *Signal acquisition device*

For the initial study and to validate the ear mold, the Bio-Amplifier in the BME lab was used for signal acquisition. Then a signal acquisition device prototype was planned to be developed using either the STM32 Bluepill or ADS1299 evaluation module or the combination of STM32 Nucleo board with ADS1299 (refer Fig. 4.5).

But due to the error in SPI communication on the ADS1299 development board, the modular prototype was unable to be developed. In parallel to developing the prototype,

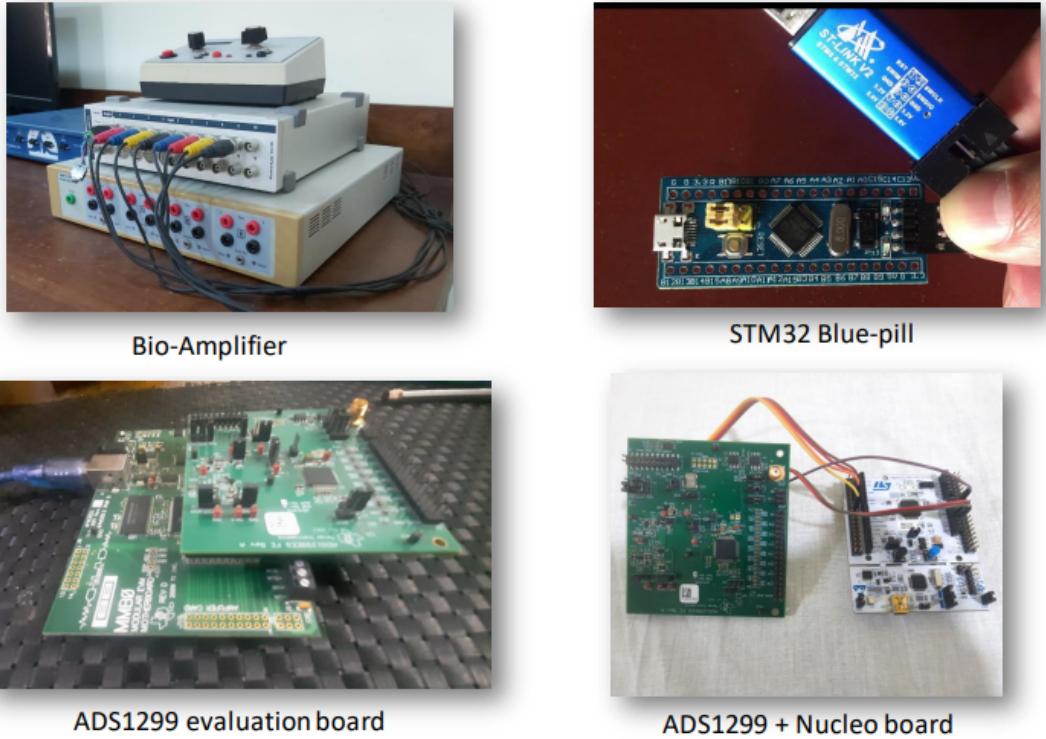


Fig. 4.5: Modules used to build the modular level signal acquisition device

the schematics of the final circuit layout was developed. Number of safety features, signal quality assurance methods and practicality constraints are included in the design phase of the schematics. They are given in the section below.

4.3.4.1 Schematics

This section gives modular level schematic diagrams designed for each of the modules for the intended hardware design such as the power supply, active shielding and filtering, analog front end, and the micro-controller. For each of the electronic modules of the ear-EEG acquisition device, a number of hazards are identified, and corresponding safety measures have been incorporated into the circuit design.

- **The Power Supply :** The power supply comprises of a 3.3V power regulator using TPS73233, 5V power regulator using TPS73250, voltage testing terminal and the power input port. Both 5.0V and 3.3V voltage inputs are needed for the analog front end ADS1299. 3.3V power supply is needed for STM32 micro-controller. For the overall safety, the supply and transmission voltage will not exceed 12V DC at any part of the circuit [47]. This supply voltage range complies with the Protective Extra Low Voltage (PELV). Therefore, even in direct skin contact, it won't pose any dangers.

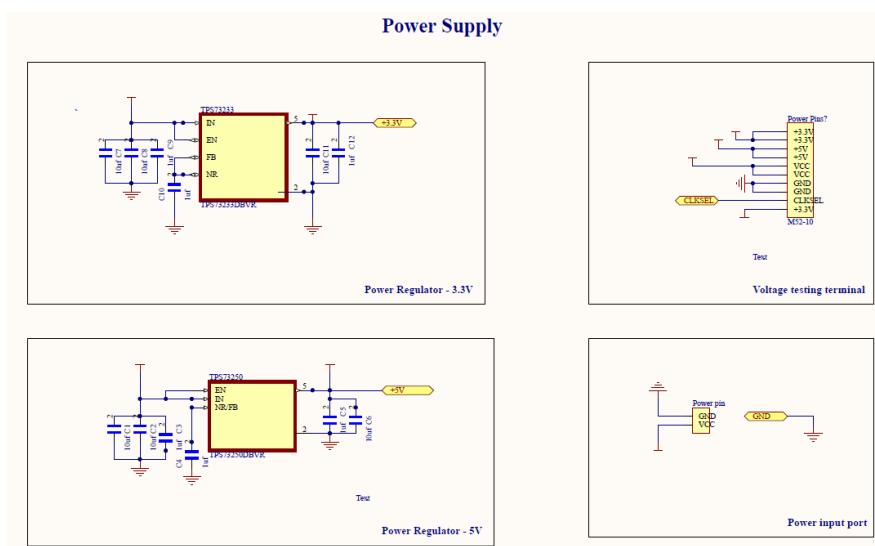


Fig. 4.6: The Power Supply

- **Active Shielding and Filtering :** A quad buffer AD8244, with high input impedance is used to isolate the active electronic circuit and the patient. This can be used to avoid overloading of the input signal and to have an active shielded circuit. The shielding is implemented on the PCB itself rather than implementing near the electrode, which is the common practice, so that all the active electronic parts are placed away from the user to avoid electric shocks and other thermal hazards.

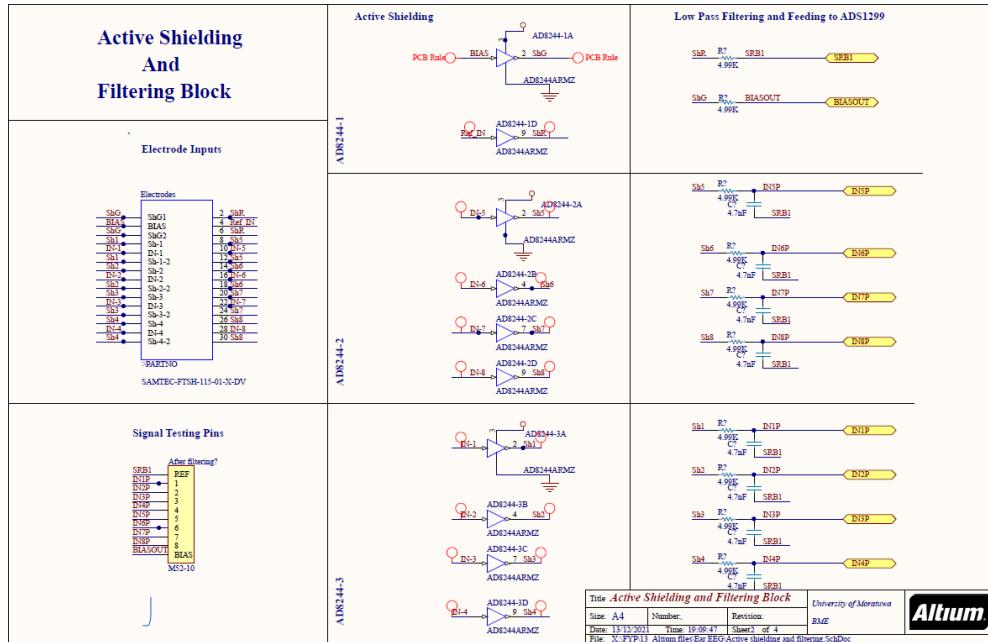


Fig. 4.7: This schematics shows the electrode input terminals, the active shielding and the passive low pass filtering blocks

- **Analog Front End :** Filtered biosignal is fed to the ADS1299 analog front end for pre-amplification using the programmable gain amplifiers. Following that the analog signal is quantized into digital signal and will be sent to the STM32 micro-controller. This component is chosen to maintain the superior signal quality during acquisition.

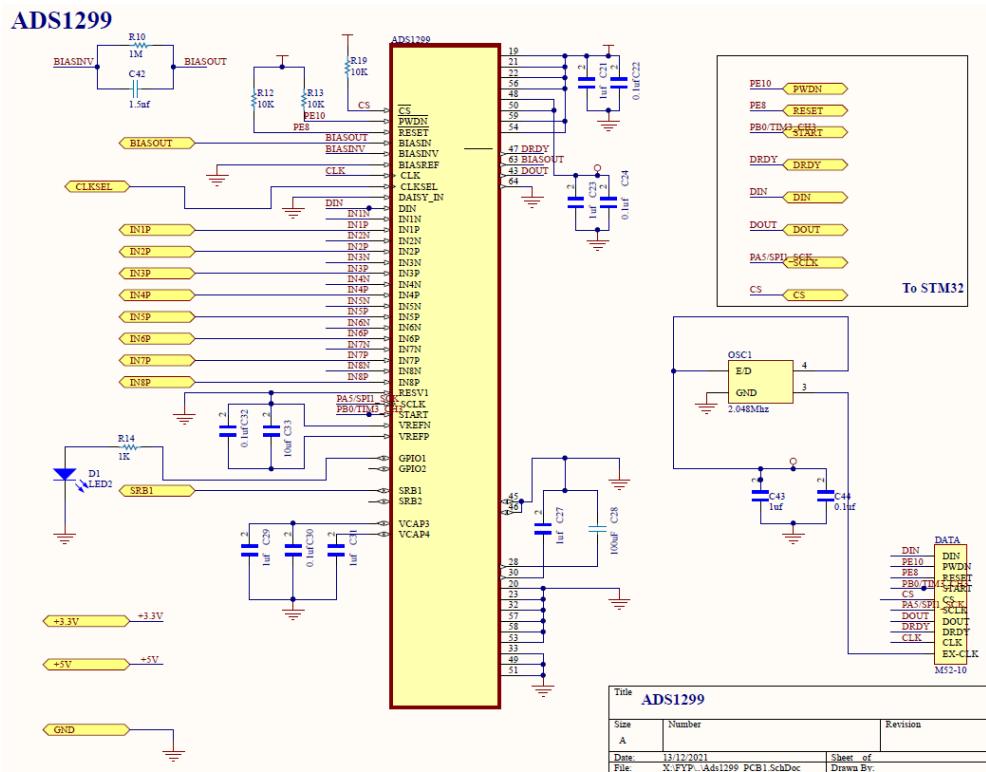


Fig. 4.8: ADS1299 - The Analog Front End Circuit

- Micro-controller :**The STM32F411 is selected as the micro-controller because of its best-in-class system performance for code execution, data transfer and signal processing. Also, it is power efficient and cost effective. For the modular hardware development, NUCLEO-F411RE development board has been used. Following that, STM32F411 micro-controller is intended to be used for the final PCB implementation of the complete circuit.

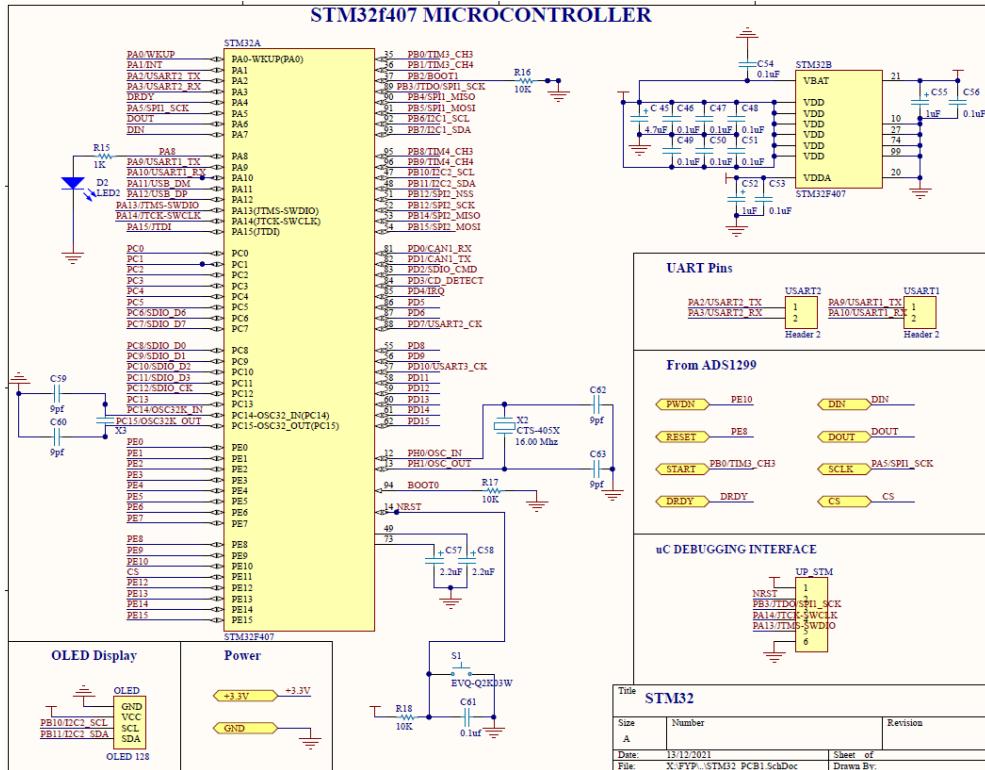


Fig. 4.9: STM32 - The Micro-controller Circuit

4.4 Challenges in Execution

As mentioned earlier, we had planned to develop a hardware setup for the signal acquisition and we went to the extent of developing the schematics. However, even though we made the necessary purchase requests for the required components during the latter part of the last year, the existing economic crisis made it impossible to receive them on time. Thus, we were forced to fall back to available alternatives. So instead of the Ag-AgCl silver pellet electrodes with which we have originally planned to acquire the ear-EEG signals, we have modified the button electrodes that were available and used them for signal acquisition. Also, we had to use the normal EEG cables instead of the shielded coaxial cables with which we have originally intended to carry out the signal transmission. The disadvantage of the employed alternative is the absence of both active and passive shielding.

Have to make AgCl electrodes

shielded coaxial

As mentioned earlier, we were also not able to use the ADS1299 board that we had access to and therefore, we have used the Bio-amplifier available in the BME lab for signal acquisition.

4.5 The Experimental procedure

The experiment was conducted in two phases. In the initial phase, the primary focus was the connectivity of the device from the electrode to signal acquisition setup, comfort to the user, manageable skin to electrode resistance levels, and visualization of the raw signal. For this purpose, the subjects ear was cleaned with alcohol and then the top layer of skin was scraped off using abrasive gel. The subject was asked to wear the ear impression with the dry electrode and wet electrode. Since there was no significant difference in the user comfort level the skin to electrode resistance become the major deciding factor in choosing the type of electrodes to be used. In the experiment, an unacceptable high margin of electrode to skin resistance from the dry electrodes was observed. On the other hand, wet electrode setup had resistance values below 5 kOhms which is in the acceptable range. Therefore, the wet electrode setup was finalized for the future mold designs. The ear mold design had one electrode in each side. The mastoid was used as the common reference point for the acquired signals. In addition EOG electrodes were placed in the temple area. The subject was asked to perform simple activities like, vertical and horizontal eye movements, blinking, and teeth clenching. The corresponding artefacts were clearly observed on the recorded raw signals. Hence, it can be concluded that, the acquired raw signal is in better quality with high SNR and responds well for EOG and EMG artefacts. In addition it was observed that for specific subjects, for specific molds a clear ECG pattern was recorded apart from the other intended signals. But this result was not observed from all the subjects who underwent the study. After a thorough analysis, it was found that, if the electrodes placed in the ear canal falls over any major artery running in the ear canal, the ECG artefact will be added to the acquired biosignal.

Following that, in the second phase, the subject was asked to sleep in the lab setup with the ear mold and EOG sensors. At this experiment, two electrodes were intro-



Fig. 4.10: Basic study procedures conducted on the subject



Fig. 4.11: Sleep study procedures conducted on the subject

duced in each side of the mold. Therefore both inter-ear and intra-ear signals can be obtained. Similar to the previous phase, EOG sensors were placed in the temple region. The subject was asked to sleep and signal was acquired until the subjects wakes up without any external influence. The subject was inquired about the quality of sleep, comfort of the ear mold. The signal acquired during this sleep study was later further analysed offline. Refer subsection [5.2](#) for the results.

4.6 Ethical Clearance

After the completion of the intended hardware setup, we require ethical clearance to collect data in order to validate and assess the quality of our device. Following ethical clearance documents were prepared and submitted to the ethical review committee of University of Moratuwa and the reviewers have suggested to make few minor modifications.

- Research Proposal
- Consent Form
- Information sheet

The revised version of the documents addressing the reviewers' concerns have been resubmitted to the ethical review committee and were approved. As of now, the target population for data collection is the healthy undergraduates of University of Moratuwa, and study location will be the Bionics Lab, Department of Mechanical Engineering, University of Moratuwa.

4.7 Future Prospects

The immediate future of the hardware aspects of this project is the completion of the intended hardware design. In addition to that we also can have a multiple ear-EEG signal acquisition setup which can increase the quality of the acquired signals. Also,

we can go for a wireless transmission design where the circuits can be placed within the enclosure that we have designed. Also, we can go for 3D printing of the ear mold design where even the conductive material can be 3D printed within the mold, which can improve the accuracy and precision of signal acquisition.

Chapter 5

RESULTS AND DISCUSSION

5.1 Cross-Modal Transformer

This section focuses on the conducted experiments and the performance by our proposed cross-modal transformer on a publicly available sleep-EDF-expanded 2018 dataset.

5.1.1 *Experiments*

We conducted experiments on the public dataset using EEG and EOG signals to validate our algorithm and compare with the state-of-the-art methods.

5.1.1.1 *Training Setup*

We defined two experimental setups based on the classification schemes (one-to-one and many-to-many) utilizing our proposed epoch and sequence cross-modal transformers. Both models are trained using Adam optimizer [48] with learning rate (lr), β_1 and β_2 set to, 10^{-3} , 0.9 and 0.999, respectively. The weight decay of 0.0001 is applied to avoid overfitting. The batch size is experimentally chosen as 32. We used weighted categorical cross entropy as the loss function (\mathcal{L}) for 5-class classification and the weights for the classes are set to $\{1, 2, 1, 2, 2\}$, to handle data imbalance. For sequence cross-modal transformer we empirically chose the number of epochs in an input sequence as $N = 5$. 5-fold cross validation is used to evaluate the performance of the model, and to tune both hyper-parameters and model architecture. The model is implemented in Pytorch environment and trained using a Nvidia Quadro RTX 5000 graphics card with 16 GB memory.

5.1.1.2 *Evaluation Metrics*

In our study we used accuracy (ACC), Cohen’s kappa coefficient (κ), macro averaged F1-score (MF1), sensitivity (Sens.), specificity (Spec.) and macro averaged G-mean (MGm) as the performance evaluation metrics of our model. Macro averaged F1-score is the average of F1-scores of all 5 classes. MF1 and MGm are the suitable performance metrics for imbalanced datasets. We have reported per class F1-scores to evaluate the models performance accurately. Given the True Positives (TP), True Negative (TN),

False Positive (FP) and False Negative (FN) the ACC, MF1, Sens., Spec. and MGm are defined as follows,

$$ACC = \frac{TP + TN}{TN + TP + FN + FP}$$

$$MF1 = 2 \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

where, $Precision = \frac{TP}{TP+FP}$.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$MGm = \sqrt{Specificity \times Recall}$$

5.1.2 Sleep Stage Classification Performance

We present the overall performance of both epoch and sequence cross-modal transformers on the sleep-EDF-expanded 2018 dataset for both classification schemes in Table 5.1 and class-wise performance in Table 5.2. Our proposed sequence cross-modal transformer achieves state-of-the-art performance in terms of sensitivity, specificity and G-mean. In comparison to those reported in previous works, sequence cross-modal transformer outperforms majority of them with the overall accuracy of 83.54%, Cohen’s Kappa of 0.774 and macro-averaged f1-score of 78.31.

Difference in performance between our sequence cross-modal transformer with the current state-of-the-art method XSleepNets [49] is very small. Compared to XSleep-Nets, our sequence cross-modal transformer takes lesser time to train as shown in Table 5.3 and was able to achieve higher performance with less number of PSG epochs in the sequence. We compared the performance with the existing transformer based method SleepTransformer [15] on sleep-EDF-expanded 2018 dataset without any pretraining on a larger database, where our sequence cross-modal transformer gave superior performance. This shows the modeling capability of our cross-modal transformer on a relatively smaller dataset compared to SleepTransformer. We strongly believe that deep learning techniques such as transfer learning, knowledge distillation, large-scale training and self-supervised pretraining could further improve the performance of our

method. Our epoch cross-modal transformer was able to achieve good performance by considering only one PSG epoch as input and with significantly less number of parameters compared to majority of the previous works.

TABLE 5.1: Performance comparison between cross-modal transformer and previous works on sleep-EDF-expanded 2018 dataset.

Method	Channels	Ep/ Seq	No of Params	Overall Performance					
				ACC	κ	MF1	Sens.	Spec.	MGm
SleepEEGNet[50]	Fpz-Cz	10	$\sim 2.6M$	80.0	0.73	73.6	—	—	—
DeepSleepNet[51]	Fpz-Cz	25	$\sim 24.7M$	77.1	0.69	71.2	—	—	—
MultitaskCNN[52]	Fpz-Cz	—	—	79.6	0.72	72.8	—	—	82.5
AttnSleep[52]	Fpz-Cz	3	—	82.9	0.77	78.1	—	—	85.6
SleepTransformer[15]	Fpz-Cz	21	$\sim 3.7M$	81.4	0.74	74.3	74.5	95.0	84.1
SeqSleepNet[51]	Fpz-Cz	20	$\sim 0.2M$	82.6	0.76	76.4	76.3	95.4	85.3
FCNN+RNN[49]	Fpz-Cz	20	$\sim 5.6M$	82.8	0.76	76.6	75.9	95.4	85.1
XSleepNet2[49]	Fpz-Cz	20	$\sim 5.8M$	84.0	0.78	77.9	77.5	95.7	86.1
XSleepNet1[49]	Fpz-Cz	20	$\sim 5.8M$	83.6	0.77	77.8	77.7	95.7	86.2
NaiveFusion[49]	Fpz-Cz	20	$\sim 5.8M$	82.3	0.76	76.2	75.7	95.3	84.9
TinySleepNet[51]	Fpz-Cz	15	$\sim 1.3M$	83.1	0.77	78.1	—	—	—
XSleepNet2[49]	Fpz-Cz EOG	20	—	84.0	0.78	78.7	77.6	95.7	86.2
XSleepNet1[49]	Fpz-Cz EOG	20	—	84.0	0.78	78.4	77.1	95.6	85.9
SeqSleepNet[49]	Fpz-Cz EOG	20	—	83.8	0.78	78.2	77.4	95.6	86.0
FCNN+RNN[49]	Fpz-Cz EOG	20	—	82.7	0.76	76.9	75.5	95.3	84.8
NaiveFusion[49]	Fpz-Cz EOG	20	—	82.5	0.76	76.9	75.8	95.3	85.0
Epoch Cross- Modal Transformer	Fpz-Cz EOG	1	$\sim 1.3M$	80.5	0.73	74.2	75.2	95.0	84.5
Sequence Cross- Modal Transformer	Fpz-Cz EOG	5	$\sim 5.6M$	83.5	0.77	78.3	79.3	95.7	87.2

When considering class-wise performance based on F1-score, our sequence cross-modal transformer achieves state-of-the-art performance in the prediction of wake, N1 and REM sleep stages. This enhancement in the performance can be attributed to the capability of our method to learn cross-modal relationships, where EOG along with EEG makes an impact on their predictions. The performance of our sequence cross-modal transformer in predicting N2 and N3 stages are on-par with the previous work.

TABLE 5.2: Class-wise performance comparison between cross-modal transformer and previous works on sleep-EDF-expanded 2018 dataset.

Method	Channels	Per-class Performance				
		W	N1	N2	N3	REM
SleepEEGNet[50]	Fpz-Cz	91.7	44.1	82.5	73.5	76.1
DeepSleepNet[51]	Fpz-Cz	90.4	46.0	79.1	68.6	71.8
MultitaskCNN[52]	Fpz-Cz	90.9	39.7	83.2	76.6	73.5
AttnSleep[52]	Fpz-Cz	92.6	47.4	85.5	83.7	81.5
SleepTransformer[15]	Fpz-Cz	91.7	40.4	84.3	77.9	77.2
SeqSleepNet[51]	Fpz-Cz	91.8	42.6	86.5	76.4	84.1
FCNN+RNN[49]	Fpz-Cz	92.5	47.3	85.0	79.2	78.9
XSleepNet2[49]	Fpz-Cz	93.3	49.9	86.0	78.7	81.8
XSleepNet1[49]	Fpz-Cz	92.6	50.2	85.9	79.2	81.3
NaiveFusion[49]	Fpz-Cz	93.2	49.6	86.2	79.4	82.5
TinySleepNet[51]	Fpz-Cz	92.8	51.0	85.3	81.1	80.3
Epoch Cross-Modal Transformer	Fpz-Cz	92.2	44.4	82.7	75.0	76.8
Sequence Cross-Modal Transformer	EOG					
	Fpz-Cz	93.3	52.4	84.8	75.6	85.5
	EOG					

5.1.3 Computational Complexity

Table 5.3 shows the comparison of model size in terms of parameters and training time take for 1000 steps between our proposed cross-modal transformers and the previously reported works. Here, our method trains faster compared to current state-of-the-art XSleepNets because their architecture is based on recurrent neural networks (RNNs) which requires to process the data in order. Unlike RNNs transformers are capable of processing data in a parallel manner. The parallelism in the transformers enabled our proposed method to train faster compared to current state-of-the-art XSleepNets and other previous works as given in Table 5.3. Our epoch transformer achieves performance closer to the SleepTransformer model with significantly smaller model print and low training time.

TABLE 5.3: Comparison of model size and training time.

Method	No of Parameters	Training time per 1000 steps (s)
SleepTransformer[15]	$\sim 3.7M$	308
XSleepNet2[15]	$\sim 5.8M$	828
SeqSleepNet[15]	$\sim 2.7M$	379
Epoch Cross-Modal Transformer	$\sim 1.3M$	53
Sequence Cross-Modal Transformer	$\sim 5.6M$	174

5.1.4 Interpretation results

Along with competitive sleep stage classification performance, the major contribution of cross-modal transformers are their interpretability. As mentioned previously, we leverage the attention mechanisms to interpret the results as intra-modal, cross-modal and inter-epoch relationships illustrated in Fig. 5.1 and Fig. 5.2. Here, we interpret the predictions of sequence cross-modal transformer for a sequence of PSG epochs.

Fig. 5.1 illustrates the prediction for a sequence of PSG epochs in N3, N3, N3, N2 and wake stages respectively. Inter-epoch attention graphs visualize the relationships between the epochs. Consider epoch 4 in the sequence, where the impact of previous epochs in the prediction is clearly visualized (According to sleep cycle a person can go to N2 sleep stage from N3 sleep stage). Cross-modal attention graphs interprets

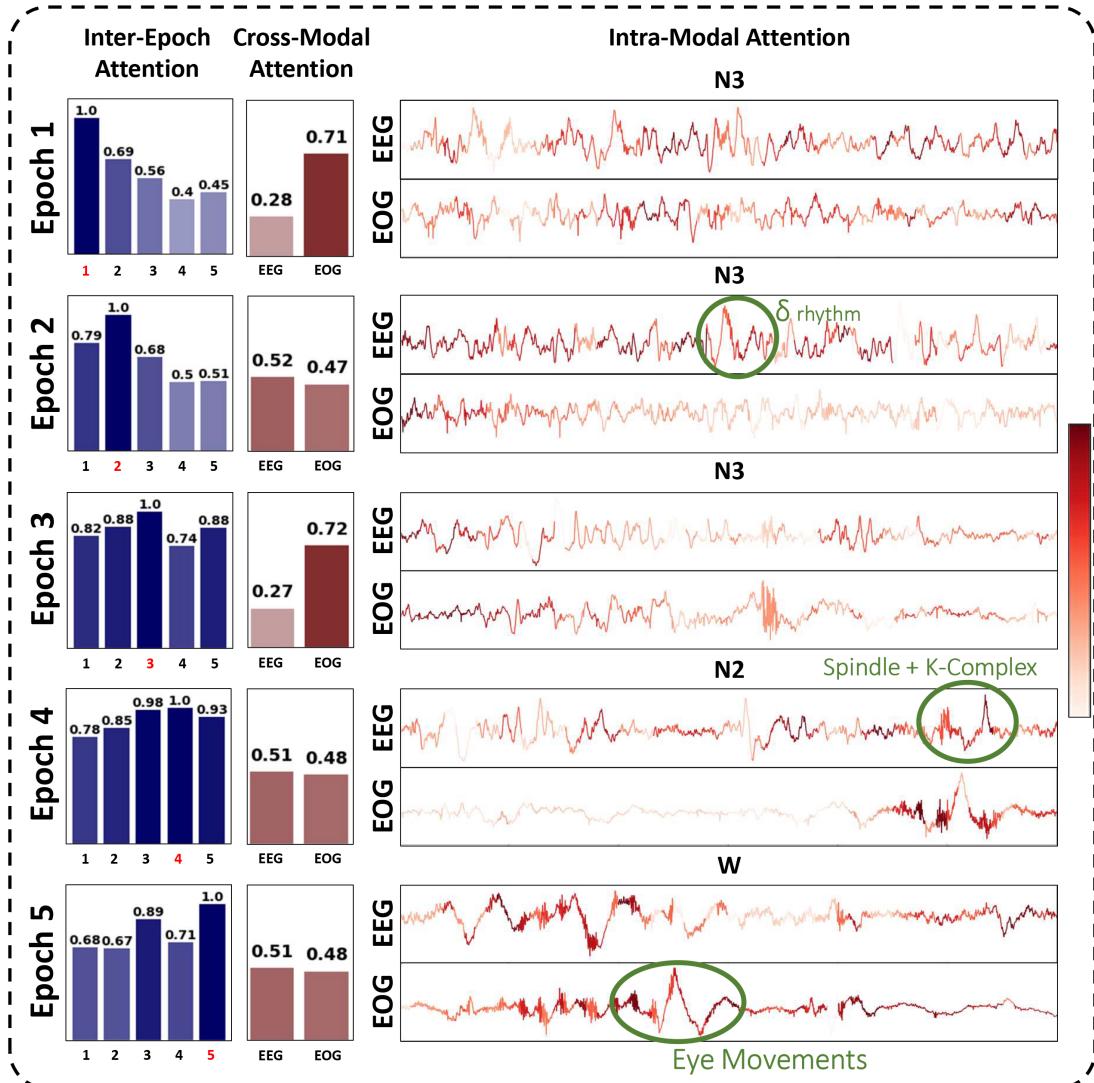


Fig. 5.1: Interpretation results for sequence of PSG epochs in N3, N3, N3, N2 and wake stages respectively.

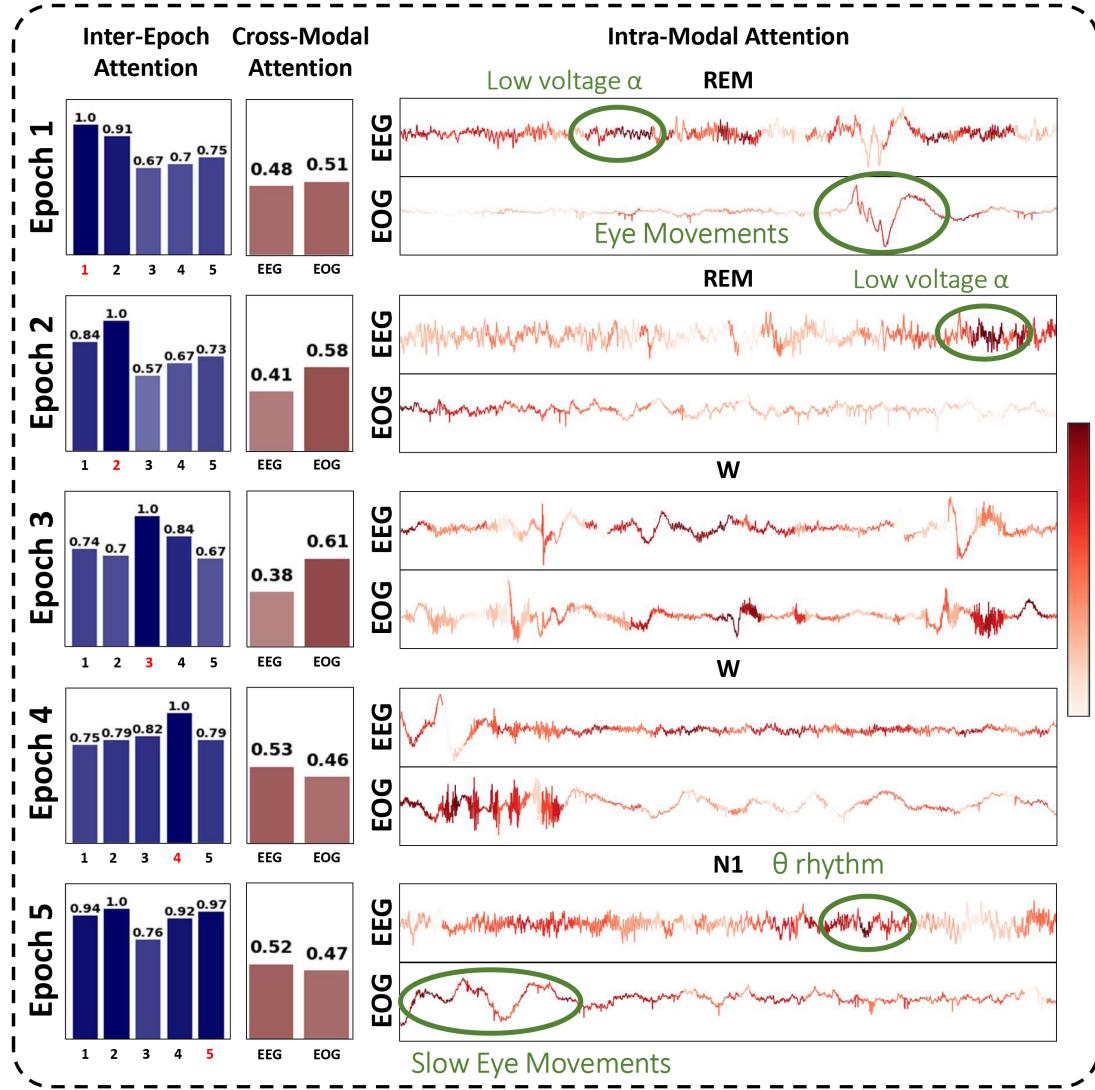


Fig. 5.2: Interpretation results for sequence of PSG epochs in REM, REM, wake, wake and N1 stages respectively.

the impact of each modality on the prediction. Finally, the intra-modal attention maps interprets, which time segments in the PSG signal impacts the predictions. The darker areas in the signals had more impact on the prediction. Our method was able to highlight important patterns in the signals corresponding to specific sleep stages, which shows the reliability of our method. In epoch 2 the δ rhythm is highlighted which is a feature of N3 sleep stage. Spindles and K-complexes occurs in N2 sleep stages, which is highlighted in epoch 4. During wake stage, muscle artifacts in the EEG and EOG channels increases and eye movements can be observed in EOG channel. These patterns are clearly highlighted in epoch 5.

Similarly, Fig. 5.2 shows the interpretations of sequence of PSG epochs in REM, REM, W, W and N1 stages respectively. In the cross-modal attention graphs for REM stages and wake stages, shows that EOG modality had more impact on prediction. This

can be attributed to the eye movements occurring in these stages, which are captured in EOG signals. The eye movements are clearly highlighted in intra-modal attention maps of epoch 1 and 5. Additionally, the method was able to highlight low voltage α rhythm, which is a pattern occurs during REM stage. We strongly believe that developing interpretable deep learning methods is the most feasible way forward to use artificial intelligence for clinical applications.

5.2 Validation of Hardware Implementation

Our final implementation of signal acquisition setup includes modified button electrodes, EEG cables and Bio-Amplifier. For this setup, we have carried out validation on the acquired signals to verify whether they demonstrate the patterns of EEG signals. For that we have conducted two kinds of studies namely, the basic study and the rigorous study. Under the basic study, the subject was asked to perform the activities of teeth clenching and eye movements and we have verified whether the corresponding EMG and EOG artifacts could be observed in the acquired signals. In addition to that we were also able to see occasional ECG signals during the signal acquisition as shown in Fig. 5.3.

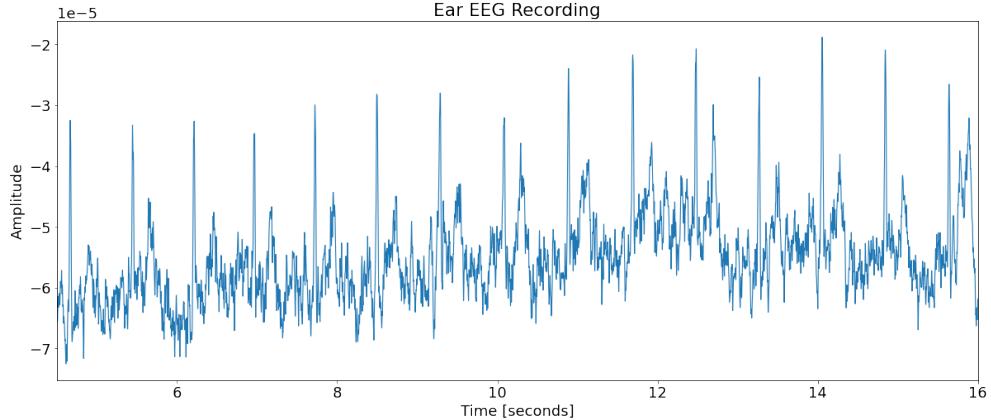


Fig. 5.3: ECG artifacts observed during ear EEG signal acquisition

Under the rigorous study, the subject was asked to sleep for a prolonged period (around one hour) and the data acquired during that period is analyzed offline. For the acquired data, we have carried out the power spectrum analysis and from the PSD plot shown in Fig. 5.4, it can be clearly seen that the power is high in the low frequency region, which is the expected behaviour during sleep.

Also, we have carried out band-wise frequency analysis as shown in Fig. 5.5 and it can be observed from the plots that in both the alpha and beta frequency bands, the power has been decreasing with time whereas in both delta and theta frequency bands we could see a gradual increment in the power. All these four observations in

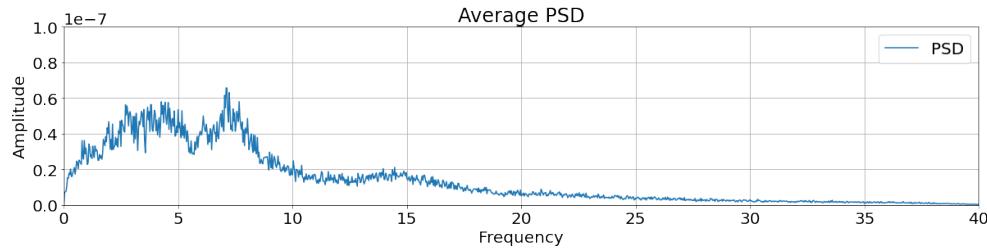


Fig. 5.4: Power Spectral Density plot of the ear EEG data acquired during the sleep study

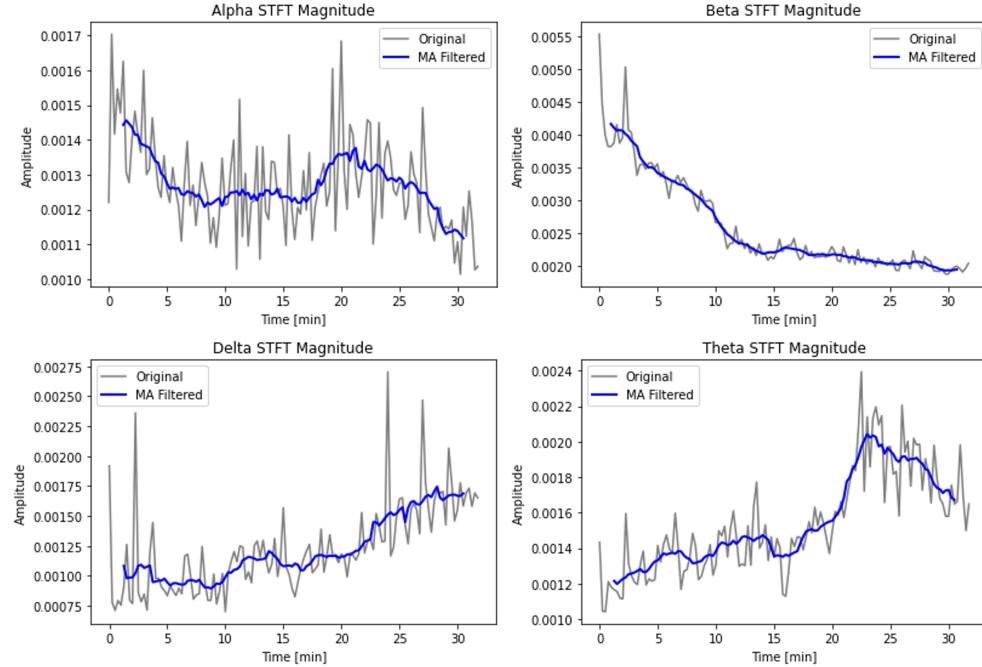


Fig. 5.5: STFT magnitude plots of alpha, beta, theta and delta frequency bands with time, for the ear EEG data acquired during the sleep study

the frequency bands clearly indicate that the subject has been transitioning from light sleep to deep sleep during the signal acquisition period.

REFERENCES

- [1] M. W. Mahowald, “Book Review Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem By the Committee on Sleep Medicine and Research,” *N. Engl. J. Med.*, Vol. 356, No. 2, 2007, pp. 199-200.
- [2] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification,” *J. Clin. Sleep Med.*, Vol. 3, No. 7. 2007.
- [3] J. Allan Hobson, “A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects,” *Electroencephalogr. Clin. Neurophysiol.*, Vol. 26, No. 6, 1969, pp. 305-310.
- [4] H. Danker-hopfeet al., “Interrater Reliability for Sleep Scoring According to The Rechtschaffen & Kales and The New AASM Standard,” *J. Sleep Res.*, Vol. 18, No. 1, 2009, pp. 74–84.
- [5] Testing your Patients Sleep Apnea at Home, Cardio Sleep Solutions, <https://cardiosleepsolutions.com/testing-your-patients-sleep-apnea-at-home/>
- [6] H. W. Loh et al., “Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020),” *Appl. Sci.*, Vol. 10, No. 24. 2020, p. 8963.
- [7] T. Zhu, W. Luo, and F. Yu, “Convolution-and attention-based neural network for automated sleep stage classification,” *Int. J. Environ. Res. Public Health*, Vol. 17, No. 11, 2020, p. 4152.
- [8] X. Zhang et al., “Automated multi-model deep neural network for sleep stage scoring with unfiltered clinical data,” *Sleep Breath.*, Vol. 24, No. 2, 2020, pp. 581-590.
- [9] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, “Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification,” *IEEE Trans. Biomed. Eng.*, Vol. 66, No. 5, 2019, pp. 1285-1296.
- [10] O. Yildirim, U. B. Baloglu, and U. R. Acharya, “A deep learning model for automated sleep stages classification using PSG signals,” *Int. J. Environ. Res. Public Health*, Vol. 16, No. 4, 2019, p. 599.

- [11] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, “SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 27, No. 3, 2019, pp. 400-410.
- [12] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, “Automatic Sleep Stage Classification Using Single-Channel EEG: Learning Sequential Features with Attention-Based Recurrent Neural Networks,” in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2018.
- [13] M. Längkvist, L. Karlsson, and A. Loutfi, “Sleep Stage Classification Using Unsupervised Feature Learning,” *Adv. Artif. Neural Syst.*, July 2012.
- [14] O. Tsinalis, P. M. Matthews, and Y. Guo, “Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders,” *Ann. Biomed. Eng.*, Vol. 44, No. 5, 2016, pp. 1587-1597.
- [15] H.P. Phan et al., “SleepTransformer: Automatic Sleep Staging with Interpretability and Uncertainty Quantification,” *IEEE. Trans. Biomed. Eng.*, 2022.
- [16] J. B. Stephansen et al., “Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy,” *Nat. Commun.*, Vol. 9, No. 1, 2018, pp. 1-15.
- [17] A. Supratak, H. Dong, C. Wu, and Y. Guo, “DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 25, No. 11, 2017, pp. 1998-2008.
- [18] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, “Mixed Neural Network Approach for Temporal Sleep Stage Classification,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 26, No. 2, 2018, pp. 324-333.
- [19] T. Nakamura, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, “Automatic sleep monitoring using ear-EEG,” *IEEE J. Transl. Eng. Health Med.*, Vol. 5, 2017, pp. 1-8.
- [20] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, “Hearables: automatic overnight sleep monitoring with standardized in-ear EEG sensor”. *IEEE Trans. Biomed. Eng.*, Vol. 67, No. 1, 2019, pp. 203-212.
- [21] K. B. Mikkelsen et al., “Accurate whole-night sleep monitoring with dry-contact ear-EEG”. *Sci. Rep.*, Vol. 9, No. 1, 2019, pp. 1-12.
- [22] J. H. Park, D. G. Jang, J. W. Park, and S. K. Youm, “Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor,” *Sensors*, Vol. 15, No. 9, 2015, pp. 23402–23417.

- [23] V. Goverdovsky et al., “Hearables: Multimodal physiological in-ear sensing,” *Sci. Rep.*, Vol. 7, No. 1, 2017, pp. 1–10.
- [24] C. Sun, C. Chen, J. Fan, W. Li, Y. Zhang, and W. Chen, “A hierarchical sequential neural network with feature fusion for sleep staging based on EOG and RR signals,” *J. Neural Eng.*, Vol. 16, No. 6, 2019, p. 066020.
- [25] Y. Wei, X. Qi, H. Wang, Z. Liu, G. Wang, and X. Yan, “A Multi-Class Automatic Sleep Staging Method Based on Long Short-Term Memory Network Using Single-Lead Electrocardiogram Signals,” *IEEE Access*, Vol. 7, 2019, pp. 85959–85970.
- [26] N. Sridhar et al., “Deep learning for automated sleep staging using instantaneous heart rate,” *NPJ Digit. Med.*, Vol. 3, No. 1, 2020, pp. 1-10.
- [27] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, “In-Ear EEG From Viscoelastic Generic Earpieces: Robust and Unobtrusive 24/7 Monitoring,” *IEEE Sens. J.*, Vol. 16, No. 1, 2016, pp. 271–277.
- [28] K. B. Mikkelsen, S. L. Kappel, D. P. Mandic, and P. Kidmose, “EEG recorded from the ear: Characterizing the Ear-EEG Method,” *Front. Neurosci.*, Vol. 9, Nov. 2015, p. 438.
- [29] A. R. Bertelsen et al., “Generic Dry-Contact Ear-EEG,” 2019 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2019.
- [30] P. Kidmose, D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic, “A study of evoked potentials from ear-EEG,” *IEEE Trans. Biomed. Eng.*, Vol. 60, No. 10, 2013, pp. 2824–2830.
- [31] A. Sterr et al., “Sleep EEG derived from behind-the-ear electrodes (cEEGrid) compared to standard polysomnography: A proof of concept study,” *Front. Hum. Neurosci.*, Vol. 12, No. November, 2018, pp. 1–9.
- [32] R. Kaveh et al., “Wireless User-Generic Ear EEG,” *IEEE Trans. Biomed. Circuits Syst.*, Vol. 14, No. 4, 2020, pp. 727–737.
- [33] O. Valentin et al., “Custom-fitted in-and around-the-ear sensors for unobtrusive and on-the-go eeg acquisitions: Development and validation,” *Sensors*, Vol. 21, No. 9, 2021, p. 2953.
- [34] S. Passler, N. Müller, and V. Senner, “In-ear pulse rate measurement: A valid alternative to heart rate derived from electrocardiography?,” *Sensors*, Vol. 19, No. 17, 2019, p. 3641.

- [35] B. Vescio, M. Salsone, A. Gambardella, and A. Quattrone, "Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings," Sensors, Vol. 18, No. 3, 2018, p. 844.
- [36] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Oberye, "Analysis of a Sleep-Dependent Neuronal Feedback Loop: the Slow-Wave Microcontinuity of the EEG," IEEE Trans. Biomed. Eng., vol. 47, No. 9, 2000, pp. 1185–1194.
- [37] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation, Vol. 101, No. 23, June 2000, pp. e215–e220.
- [38] A. Vaswani et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst., Vol. 30, 2017.
- [39] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. In-man, "1D Convolutional Neural Networks and Applications: A Survey," Mech. Syst. Signal Process, Vol. 151, 2021, p. 107398.
- [40] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," Circulation, Vol. 101, No. 23, 2000, pp. e215–e220.
- [41] A. Rechtschaffen, "A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects," Brain information service, Brain Research Institute, US Dept. of Health, Education, and Welfare, 1968.
- [42] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification," J. Clin. Sleep Med., Vol. 3, No. 7. 2007.
- [43] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," ICLR, 2021.
- [44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL), 2019, pp. 4171–4186.
- [45] K. B. Mikkelsen, D. B. Villadsen, M. Otto, and P. Kidmose, "Automatic sleep staging using ear-EEG," Biomed. Eng. Online, Vol. 16, No. 1, 2017, pp. 1-15.
- [46] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531 2, No. 7, 2015.

- [47] IEC 60050 - International Electrotechnical Vocabulary - Details for IEV number 826-12-32: "PELV system". URL: <https://www.electropedia.org/iev/iev.nsf/display?openform&ievref=826-12-32> (visited on 09/06/2021).
- [48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," ICLR, 2015, pp. 1–15.
- [49] H. Phan, O. Y. Chen, M. C. Tran, P. Koch, A. Mertins, and M. De Vos, "Xsleepnet: Multi-view sequential model for automatic sleep staging," IEEE Trans. Pattern Anal. Mach. Intell., 2021.
- [50] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated Sleep Stage Scoring with Sequence to Sequence Deep Learning Approach," PloS one, Vol. 14, No. 5, 2019, p. e0216456.
- [51] L. Fiorillo, P. Favaro, and F. D. Faraci, "Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates," IEEE Trans. Neural Syst. Rehabilitation Eng., Vol. 29, 2021, pp. 2076–2085.
- [52] Eldele et al., "An Attention-Based Deep Learning Approach for Sleep Stage Classification with Single-Channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 29, 2021, pp. 809–818.

APPENDICES

To our Parents, Lecturers, Friends
and
the People of the Democratic Socialist Republic of Sri Lanka