

Deep Learning-Based Sleep Stage Classification Using EEG

Mehdi Shah Poori Arani

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical and Electronic Engineering

Eastern Mediterranean University
February 2023
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Assoc. Prof. Dr. Rasime Uyguroğlu
Chair, Department of Electrical and
Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Dr. Noushin Hajarolasvad
Co-Supervisor

Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

1. Prof. Dr. Hasan Demirel

2. Prof. Dr. Önsen Toygar

3. Assoc. Prof. Dr. Kamil Yurtkan

ABSTRACT

Classification of sleep stages is an essential area of research that helps develop treatments for people with sleep disorders. According to common sleep stage criteria, sleep is divided into six different stages: Wakeful sleep (W), REM (rapid eye movement) sleep, and non-REM sleep (S1-S4). Sleep processing can be performed by analyzing electroencephalogram (EEG) signals in a 30-second cycle (epoch). These stages are chosen and established on an analysis of brain workouts during sleep. This reveals a clear pattern that characterizes each stage. Sleep deprivation can cause various illnesses, including obesity, heart disease, diabetes, and reduced life expectancy [2]. Sleep professionals usually classify sleep stages into polysomnography (PSG) signals. Polysomnography consists of an electroencephalogram (EEG), electro-oculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) [2]. In addition, one category of such classifiers, Deep Learning (DL) based EEG signal classification, is used to classify sleep stages. The treatise includes an analysis of the performance of the considered methods of sleep grading. In addition, the strengths and weaknesses of classical and deep learning-based sleep staging methods will be explored. In addition, we compared standard classification with the data fusion methods with their accuracy.

Keywords: S1 - 4, W (Wake), Sleep Stage, Deep Learning, Accuracy, Data Fusion

ÖZ

Uyku aşamalarının sınıflandırılması, uyku bozukluğu olan kişiler için tedavilerin geliştirilmesine yardımcı olan önemli bir araştırma alanıdır. Yaygın uyku evresi kriterlerine göre uyku altı farklı evreye ayrılır: uyanık uyku (W), REM (hızlı göz hareketi) uykusu ve REM olmayan uyku (S1-S4). Uyku işleme, bir elektroensefalogram (EEG) analiz edilerek gerçekleştirilebilir. 30 saniyelik bir döngüde (dönem) sinyaller. Bu aşamalar, uyku sırasındaki beyin egzersizlerinin analizine göre seçilir ve kurulur. Bu, her aşamayı karakterize eden net bir modeli ortaya çıkarır. Uyku yoksunluğu, obezite, kalp hastalığı ve diyabet gibi çeşitli hastalıklara neden olarak yaşam beklentisini azaltabilir [2]. Uyku uzmanları genellikle uyku aşamalarını polisomnografi (PSG) olarak sınıflandırır. Polisomnografi bir elektroensefalogram (EEG), electro-oculogram (EOG), elektromiyogram (EMG) ve elektrokardiyogramdan (EKG) oluşur [2]. Bu görev, uyku aşamalarını sınıflandırmak için EEG sinyallerini kullanır. Sinyal işleme teknikleri, Veri birleştirme yöntemi gibi standart sınıflandırıcılarda gerekli işlevselliği çıkarır. Ayrıca uyku evrelerini sınıflandırmak için son teknoloji ürünü derin öğrenme tabanlı EEG sinyal sınıflandırması kullanılmaktadır. Tez, dikkate alınan uyku derecelendirme yöntemlerinin performansının bir analizini içerir. Ayrıca klasik ve derin öğrenmeye dayalı uyku evreleme yöntemlerinin güçlü ve zayıf yönleri keşfedilecektir. Ek olarak, standart sınıflandırma ile veri birleştirme yöntemlerini doğrulukları açısından karşılaştırdık.

Anahtar Kelimeler: S1 - 4, W (Uyanma), Uyku Aşaması, Derin Öğrenme, Doğruluk, Veri Füzyon

DEDICATION

To My Family

ACKNOWLEDGMENT

Firstly, I'd like to express my thanks to my patient and supportive supervisor, Prof.Dr.Hasan Demirel and co-supervisor Dr.Noushin Hajarolasvadi, who have supported me throughout this research project. I am extremely grateful for the friendly chats of our meetings and your support in my academic endeavors.

Finally, my biggest thanks to my family for all the support you have shown me through this research, the culmination of Two years of learning. Thanks for your support, without which I would have stopped these studies long ago.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZ.....	iv
DEDICATION.....	v
ACKNOWLEDGMENT.....	vi
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xii
1 INTRODUCTION.....	1
1.1 Sleep.....	1
1.2 An Overview of Sleep Stage Classification.....	2
1.3 Problem Definition and Thesis Objectives.....	4
1.4 Contribution.....	5
2 BACKGROUND THEORY.....	6
2.1 Introduction to neural Network	6
2.2 Deep Learning approach	6
2.3 Convolution Neural Network (CNN)	7
2.3.1 Types of Convolution Neural Network models.....	7
2.3.2 Training Convolution Nural Network	8
2.4 Loss Function	10
3 METHODOLOGY ...	12
3.1 Data Acquisition	12
3.2 Methode	12
3.3 Resnet Model.....	14
3.4 Convolution Architectures for EEG in the Literature	16

3.5 Experimental Methodology	16
3.6 Architecture Description	16
3.7 Architecture Details	17
3.7.1 Spectrogram	17
3.7.2 Nural Architecture	20
3.8 Data Pre-Processing	21
3.9 Training Methodology	21
3.10 Expectation	22
3.11 Fusion	22
4 RESULTS AND DISCUSSION.....	24
4.1 The SetConsidered by the Classifier	24
4.2 Performance	25
4.3 Score Fusion Performance	43
4.4 Final Concolusion	43
4.5 Comparise Performance	46
5 DISCUSSION AND CONCLUSION	47
5.1 Summery of Contribution	47
5.2 Observations on Classifier Performance	47
5.3 Future Direction	48
5.4 Methodological Improvements	48
REFERENCES	50

LIST OF TABLES

Table 1.1: Summary of EEG, EOG, and EMG patterns for different sleep stages according to the AASM manual	2
Table 1.2: Percentages of sleep stages in healthy adults	4
Table 3.1: Percentages of sleep stages in healthy adults	18
Table 3.2: Number of images used for CNN	21
Table 4.1: Confusion matrix with the accuracy of Mix signal (Fpz-Cz channel).....	26
Table 4.2: Average precision, recall, accuracy, and F1-score for Mix signal (Fpz-Cz channel)	27
Table 4.3: Average precision, recall, accuracy, and F1-score for Mix signal (Fpz-Cz channel).....	28
Table 4.4: Average precision, recall, accuracy, and F1-score for Mix signal (Pz-Oz channel).....	28
Table 4.5: Confusion matrix with the accuracy of Delta signal (Fpz-Cz channel)...	30
Table 4.6: Average precision, recall, accuracy and F1-score for Mix signal (Fpz-Cz channel)	30
Table 4.7: Average precision, recall, accuracy and F1-score for Mix signal (Fpz-Cz channel)	31
Table 4.8: Average precision, recall, accuracy and F1-score for Delta signal (Pz-Oz channel)	32
Table 4.9: Average precision, recall, accuracy and F1-score for Delta signal (Pz-Oz channel)	33
Table 4.10: Average precision, recall, accuracy and F1-score for Tehta signal (Fpz-Cz channel)	34

Table 4.11: Average precision, recall, accuracy and F1-score for Tehta signal (Fpz-Cz channel)	35
Table 4.12: Average precision, recall, accuracy and F1-score for Theta signal (Pz-Oz channel)	35
Table 4.13: Average precision, recall, accuracy and F1-score for Theta signal (Pz-Oz channel)	37
Table 4.14: Average precision, recall, accuracy and F1-score for Alpha signal (Fpz-Cz channel).....	37
Table 4.15: Confusion matrix with accuracy of Alpha signal (Pz-Oz channel).....	38
Table 4.16: Average precision, recall, accuracy and F1-score for Alpha signal (Pz-Oz channel).....	39
Table 4.17: Confusion matrix with accuracy of Beta signal (Fpz-Cz channel).....	40
Table 4.18: Average precision, recall, accuracy and F1-score for Beta signal (Fpz-Cz channel).....	41
Table 4.19: Average precision, recall, accuracy and F1-score for Beta signal (Fpz-Cz channel).....	42
Table 4.20: Average precision, recall, accuracy and F1-score for Beta signal (Pz-Oz channel).....	42
Table 4.21: Confusion matrix with accuracy of Mix signal (combined channel).....	43
Table 4.22: Average precision, recall, accuracy and F1-score for Mix signal (Combined channel).....	43
Table 4.23: Average precision, recall, accuracy and F1-score for Mix signal (Combined channel).....	43
Table 4.24: Average precision, recall, accuracy and F1-score for Delta signal (Combined channel).....	44

Table 4.25: Confusion matrix with accuracy of Theta signal (combined channel)....	44
Table 4.26: These values are for average precision, recall, accuracy and F1-score for Delta signal (Combined channel).....	44
Table 4.27: Confusion matrix with accuracy of Alpha signal (combined channel)....	44
Table 4.28: Average precision, recall, accuracy and F1-score for Alpha signal (Combined channel).....	45
Table 4.29: Average precision, recall, accuracy and F1-score for Alpha signal (Combined channel).....	45
Table 4.30: Average precision, recall, accuracy and F1-score for Beta signal (Combined channel).....	45
Table 4.31: Total accuracy for each signal	46
Table 4.32 Comparison table of proposed	46

LIST OF FIGURES

Figure 2.1: CNN architecture.....	10
Figure 3.1: Resnet-Related Content	15
Figure 3.2: Spectrogrames	18
Figure 3.3: Resnet 18 block diagram	22
Figure 4.1: Training plot (Fpz-Cz channel Mix Signal)	26
Figure 4.2 Loss(Fpz-Cz channel Mix Signal).....	26
Figure 4.3: Training plot (Pz-Oz channel Mix Signal)	27
Figure 4.4: Loss(Pz-Oz channel Mix Signal).....	28
Figure 4.5: Training plot (Fpz-Cz channel Delta Signal)	29
Figure 4.6: Loss(Fpz-Cz channel Delta Signal).....	30
Figure 4.7: Training plot (Pz-Oz channel Delta Signal)	31
Figure 4.8: Loss(Pz-Oz channel Delta Signal).....	31
Figure 4.9: Training plot (Fpz-Cz channel Theta Signal)	33
Figure 4.10: Loss(Fpz-Cz channel Theta Signal).....	33
Figure 4.11: Training plot (Pz-Oz channel Theta Signal)	34
Figure 4.12: Loss(Pz-Oz channel Theta Signal).....	35
Figure 4.13: Training plot (Fpz-Cz channel Alpha Signal)	36
Figure 4.14: Loss(Fpz-Cz channel Alpha Signal).....	37
Figure 4.15: Training plot (Pz-Oz Alpha Mix Signal)	38
Figure 4.16: Loss(Pz-Oz channel Alpha Signal).....	38
Figure 4.17: Training plot (Fpz-Cz channel Beta Signal)	40
Figure 4.18: Loss(Fpz-Cz channel Beta Signal).....	40
Figure 4.19: Training plot (Pz-Oz channel Beta Signal)	41

Figure 4.20: Loss(Pz-Oz channel Beta Signal)..... 42

Chapter 1

INTRODUCTION

1.1 Sleep

Sleep is intently associated with human health. Effective sleep quality detection helps sleep professionals monitor and test sleep disorders and formulate appropriate treatments for their patients. A scientific measure of sleep quality is polysomnography (PSG) (i.e., sleep consideration). Signals that record the activity of different parts of the human body. These collected signals mainly consist of an electroencephalogram (EEG), an electrooculogram (EOG), an electromyogram (EMG), an electrocardiogram (ECG). At PSG, a typical 8-hour sleep polysomnogram is divided into 30-second epochs. Sleep epochs are annotated into different sleep stages by technicians according to the specific rules of the sleep manual. Consistency in the rules laid out in sleep manuals is essential for sleep assessment. This is because every slight difference can lead to different annotations. To keep unity, a standard manual of rules has been considered by experts. Rechtschaffen and Kales Standard (R&K Handbook) [1] and the American Academy of Sleep Medicine (AASM) Manual [2] are two of the most widely used sleep staging manuals.

Sleep is distinguished into 6 (or 5) stages, which are Wake (W), Stage 1 (S1), Stage 2 (S2), Stage 3 (S3), Stage 4 (S4), and rapid Eye Movement (REM or R) (that is, the R&K manual [1] where S3 is further classified into S3 and S4).

Each stage is based on frequency-domain and time-domain patterns of the manual's characteristics. Summary of these scoring rules for specific sleep stages is shown in Table 1.1. Sleep evaluation is done manually by sleep experts, which is time-consuming and takes time. To remedy this, automated sleep assessment approaches have been proposed. Feature analysis can classify sleep stages automatically by feature extraction using machine learning classification algorithms.

Table 1.1: Summary of EEG, EOG, and EMG patterns for different sleep stages according to the AASM manual [2]

Stages	EEG					EOG	EMG
	Delta (<4Hz)	Theta (4-7Hz)	Alpha (8-13Hz)	Beta (>13Hz)	Time-domain patterns		
Wake			*	*		0.5-2Hz	Variable amplitude but usually higher than during sleep stages
S1		*	*		Vertex waves	Slow eye Movement	Lower amplitude than in stage Wake
S2		*			K-complexes Sleep spindles	Usually no eye movement, but slow eye movements may persist	Lower amplitude than in stage Wake and may be as low as in stage REM
S3 & S4	*				Sleep Spindles may persist.	Eye movements are not typically seen.	Lower amplitude than in stage N2 and sometimes as low as in stage REM
REM or R		*	*		Sawtooth waves	Rapid eye movement	Lower chin EMG tone; usually the lowest level of the entire recording

1.2 An Overview of Sleep Stage Classification

In 1968, a committee co-chaired by Rechtschaffen and Kales (aka R&K), divided sleep into four phases in a publication titled "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subject" [3]. Non-REM includes four classes: S1, S2, S3, and S4.

Rapid-Eye Moving stage (REM) and Wakefulness (W). Non-REM sleep itself accounts for almost 75% of sleep. The stages of sleep are described in detail below. In this context, EEG signals should be split into 30 seconds cycles.

Stage Wake (W): Corresponds to the awake state. Alpha wave activity and low voltage mixed frequency EEG are mainly present at this stage. EOG treble EMG and winks are usually found at this stage.

Stage 1 (S1): This is the bridge between waking up and falling asleep (sleep onset), also known as the light sleep stage. The heart rate begins to slow at this stage, and breathing gradually becomes regular. This stage lasts 5 to 10 minutes, yet the subject can quickly wake up. With a slow eye movement (SEM) called the sinusoidal eye movement, low-amplitude waves with frequencies between 4 and 7 Hz stand out, including peak waves with fewer than 0.5 seconds duration.

Stage 2 (S2): It's called the light sleep stage, but it's more challenging to wake up. It lowers blood pressure and temperature, reduces heart rate, and prepares the body for deep sleep. Epoch stage 2 can be evaluated by observing K-complex and sleep spindles.

Stage 3 (S3): This stage is known as the deep sleep, Delta, or Slow Wave Stage (SWS); compared to levels 1 and 2, it is more difficult to wake up the subject. During the deep stages of sleep, the body repairs and reconstructs tissues and relaxes muscles. Delta EEG with a frequency of 0.5-4 Hz can be detected with a small number of spindles (compared to level 2). If SWS occupies more than 20% of the epoch, the epoch can be considered Stage 3.

Stage 4 (S4): Accounts for about 10% of total sleep time in adults. Stage 4 of Non-REM sleep has the highest wakefulness threshold with external stimuli during different sleep stages. Arousal disorders such as sleep anxiety and sleepwalking occur during stages 3 and 4 of sleep. The amount of Non-REM stages 3 and 4 and the EEG amplitude of delta waves increase in adolescents and decrease in the elderly. This stage is strong in the first half of sleep.

REM: The main features of the REM stage are relatively low voltage, mixed frequency EEG activity, and temporary REM concurrency. The EEG pattern is close to the one described in Stage 1, distinguished by the absence of sharp mountain waves. In REM sleep, prominent sawtooth waves occur in the parietal and forehead areas, combined with rapid eye movements. Alpha wave activity is apparent at the REM stage, with a frequency 1-2 Hz slower than when awake. The Sleep Spindle and K-complex are absent, as in stage 1. Table 1.2 compares the percentage of sleep stages in healthy adults.

Table 1.2: Percentages of sleep stages in healthy adults[2]

Sleep stages	Percentage of total Sleep Time
S1	2-5%
S2	45-55%
S3/4	5-20%
REM or R	20-25%

1.3 Problem Definition and Thesis Objectives

Usually, a trained medical professional is required to classify brain waves into sleep stages which is time-consuming. This work aims to develop a fully automated machine-learning process for generating hypnograms from EEG signals that is useful

for clinical practice. Such methods should be robust to noisy data and variation between patients, technicians, and recorders. The algorithm should work well with data from various technicians and patients with abnormal sleep patterns.

Data from new patients never seen in training. Sleep stage methods should not rely on expert preprocessing or feature extraction to achieve this goal. It is important that the algorithm learns from human labels and EEG signals. Such procedure allows retraining without adjusting for new data sources or patients with specific diseases. The procedure must be able to perform on par with an expert capability.

We will go into more detail in a later section and give an example of an expert function later. Machine learning algorithms need to provide good estimates.

1.4 Contribution

A considerable number of researchers from different machine-learning fields addressed the task of automated sleep staging using EEG data.

However, none of the methods satisfies all the stated objectives to this end, this thesis employs a deep learning architecture that uses a Convolutional Neural Network (CNN) with EEG spectrogram data as input; the contribution of this work includes using colored images called spectrograms as the input data of the network. These images represent frequency changes in the time domain using pseudo colors.

Chapter 2

BACKGROUND THEORY

2.1 Introduction to Neural Networks

This chapter intends to explain the deep learning approach.

2.2 Deep Learning Approach

Artificial Neural Networks (ANNs) and Deep Learning (DL) are subsets of machine learning. However, neural networks have improved the sleep classification problem and their prediction accuracy. One of the first classified methods used in neural networks was published in 1999, to the authors' knowledge [26].

Classification problems usually use Fourier Transform (FT) methods for feature extraction. Sleep data was used from the Physionet database, and two channels were chosen for feature extraction namely, Fpz-Cz channel and Pz-Oz channel. Feature vectors from the wavelet coefficients signals time-frequency analysis (wavelet transform) for EEG classification used six stages (wake, S1-4, REM). Moreover, creating a feature to classify sleep stages is necessary and time-consuming. First, the real value EEG signals are converted into a complex number, and the converted signal is provided into the network. It is a multi-layer perceptron with special features topology and multiple hidden layers. Also, point out the problem of low classification accuracy with another approach, which also uses CNN [27]. In this study, the PSG of the Sleep-EDFx database was used, and a 6-level classification was performed.

Frequency domain information of EEG signals makes the 2D signal and creates a spectrogram fed into a CNN, where the feature extraction process is done automatically.

2.3 Convolutional Neural Networks (CNN)

CNN is inspired by the visual cortex system [22], [23]. Each neuron responds only to a limited local area in the visual cortex. Their connectivity network makes correlations embedded for input images with their classifiers. Inspired by biological and psychological research [24],[25] proposed a CNN structure trained with a backpropagation algorithm. The model is used for various recognition tasks, such as reading postal codes and handwriting recognition. CNNs' structure consists of a series of stacked layers, each layer containing such as neuron layers that receive information data through neurons and transform them into input to another neuron using a differentiable function. The three main types of CNN layers are convolutional layers, pooling layers, and fully connected (FC) layers. Layers typically run with either an activation function or a loss function transformation to adjacent layers of the network while various features are extracted from the layers. CNN has multiple applications, including image classification, object recognition, processing, and medical image analysis. It can also be used in self-driving cars and robotics and treat visual impairments; in addition, CNN is to get specific features and their combinations with higher levels. However, due to its layered architecture, it is computationally intensive, and it takes days to train such a network with large datasets; it is effective for visual tasks and exceeds all conventional methods. [14]

2.3.1 Types of Convolutional Neural Network models

Researchers proposed various types of CNNs, namely, LeNet, AlexNet, ResNet, VGGNet [20], ZfNet (2013), GoogleNet [21], XceptionNet

2.3.2 Training Convolution Neural Network

Supervised deep neural network training is formulated to minimize the loss function, and related to this, supervised deep neural networks for the search were introduced. One of the optimization strategies is the set of network parameter (or weight) values for which the loss function has a minimum value and gradient descent errors are minimized by computing gradients and updating network parameter values. The most popular and successful DL algorithm is gradient descent using a backpropagation approach where the error is backward propagated from the last layer to the first. All weights are initialized with random or probability distributions in this learning method. Inputs are routed through the network to get outcomes. Then, using the outcome and the desired output, estimate the error utilizing the cost(error) function. To understand how backpropagation works, a small CNN model is shown in Figure 2.1.

Convolutional Layers: convolution layer accepts an input matrix with a fixed number of feature maps. Mathematically, such an operation can be given as

$$Conv^{l+1} = (W^l \otimes x^l + b^l) \quad (2.1)$$

Where $Conv^k$ represents the 6 outcome feature maps of convolutional layer Conv1 and x^l denote the input for the l-th layer by x^l . Also, let $\{W^l, b^l\}$ be the weights and bias of filters in that layer. Then, the output feature map of x^l can be computed. Non-linear operations in deep networks are known as the most activation function. The following formula expresses Rectified Linear Unit (ReLU), a non-linear activation function:

$$\sigma = \max(0, x) \quad (2.2)$$

The max pooling layer is the next layer. The results of the convolutional layers are fed into the max pooling layer. A max-pooling layer takes a feature map and denote $n \times n$ squares of the feature map input by $p_{ij}^{n \times n}$ where the tuple (i, j) means the patch is

centered at location (i, j) in the input feature map. Then, the output of the sub-sampling pooling operation performs a max-pooling operation as follows:

$$p = \beta \sum_{hk} (p_{ij}^{n \times n}) + b \quad (2.3)$$

Another convolution layer is stacked on top of the previous ones to build the model's architecture shown in Figure 2.1.

Fully connected layer (FC): Once all neurons of the previous layer are connected to all neurons of the current one, we have a fully connected layer. Such layers can perform classification and is mathematically defined as

$$F^k = \sum_i w_i^k * Conv^i \quad (2.4)$$

The activation function used for classification in last layer is the softmax function

$$Z^k = softmax(F^k) = \frac{e^{F^k}}{\sum_{i=1}^6 e^{F^k}} \quad (2.5)$$

The softmax activation function produces the class labels.

All outputs sum to 1 and is in the range [0, 1]. Each output represents a probability of inputs belonging to a particular class.

Where Z^k is a vector of size equal to the number of classes.

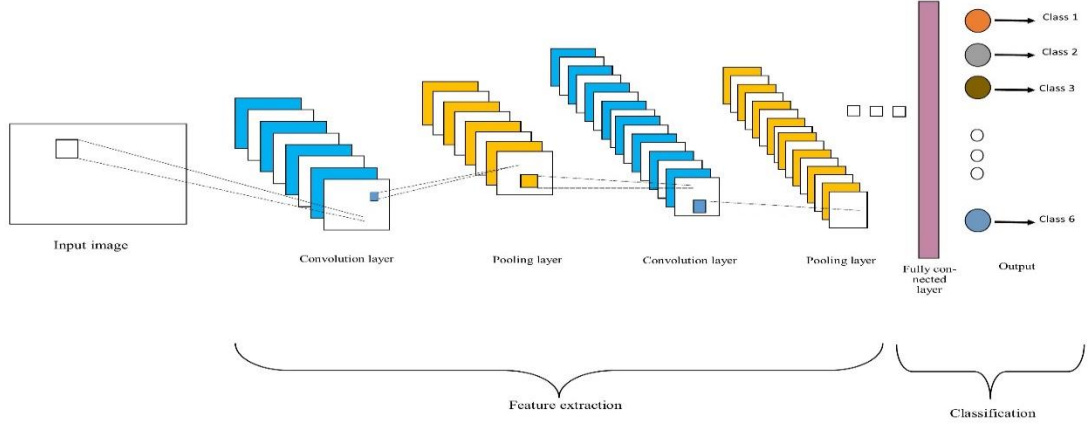


Figure 2.1: CNN architecture [15]

2.4 Loss Function

During training, inputs are fed to the network, and outputs are generated.

Calculating the error or loss is based on comparing the output to the actual label. The computed errors are used to update the network weights.

The function to compute error or loss is called the cost/loss function, and several loss functions are available. The MSE-loss function is one of the functions used, and it is not well suited for a classification problem because it assumes a regression problem with values from $-\infty$ to ∞ , which is not the case for classification problems [28]. Our purpose is to minimize the loss function between the predicted and target outputs when training. In the beginning, the Softmax function has to be defined. The Softmax function transforms the output of the different neurons into a probability distribution. Generating vector of “n” predictions from the following samples “n” data points for all variables and “y” is the vector of monitoring values of the forecasters, with “ $\hat{y}$ ” being the expected values, the predictors in model MSE calculating as

$$MSN = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.6)$$

The second function is binary Cross-Entropy Loss is a classification problem; one can apply the idea of the cross-entropy loss to a single output neuron. Again, we have to transform the output y into a probability. But because there is only one neuron, we must apply the Sigmoid function.

$$sig(y) = \frac{1}{1+e^{-y}} \quad (2.7)$$

Chapter 3

METHODOLOGY

3.1 Data Acquisition

Many approaches have been proposed to automate the sleep stage classification process.

However, these approaches have drawbacks that we try to address in this thesis:

The data used in this thesis is available on physionet.org [16]; we employed the extended version of the sleep-EDF database in the conducted experiments. There are two types of PSG recording files:

a) sleep cassette (SC), and b) sleep telemetry (ST), which belong to two different studies. The EMG data for the first study is zero amplitude. So only ST files are selected from a collection of 50 PSG signals. The EEG signal was sampled at 100 Hz, and the event marking at 1 Hz. Except for slight difficulty falling asleep, subjects were healthy without taking any sleep-related medications. The data was split into cycles of 30 seconds, and all epochs were evaluated according to R&K guidelines. These recordings include EEG (Fpz-Cz channel and Pz-Oz channel).

3.2 Method

In this section, we describe materials and methods, including different CNN models we used.

This thesis used Resnet18, which is a CNN-based classification model. We will explain its architecture further.

It is important to note that spectrograms representing a signal in frequency-time space are used as the input data of this model.

Most studies in this category were designed based on single-channel EEG.

CNNs are most commonly used for single-channel sleep classification. Deep learning architectures like CNN are usually used to extract time-invariant Features of the current sleep epoch [10], [4], [6]; then sleep signal is classified using features extracted from fully connected layers of the CNN. Some studies [5] and [8] also directly extract time-invariant features from the current sleep epoch.

Such architectures are the basic component of almost all single-channel sleep evaluation models. Additional contributing models were developed to extract target-specific features more accurately to improve deep sleep scoring performance. For example, according to the AASM manual [2], the EEG signal consists of two types of features: EEG signals and time-domain patterns that usually appear everywhere. 0.5 second; examples are K-complex and sleep spindles [10].

A smaller filter extracts time-domain patterns, and a larger filter extracts frequency-domain information from EEG signals. Moreover, different combinations of 0.5-second patterns can occur during the same sleep stage, which makes feature extraction difficult. This is because it can increase the complexity of the function at deeper layers of CNN [7], [6], [8]. Adopting a very small filter size, we extract complex time-invariant patterns from sleep epochs. Given the similarities between sleep scoring

methods and machine translation ones mentioned by Mousavi et al. [4], employing attention-based models provides a robust learning system. The attention module learns critical parts of the sleep sequence. Furthermore, the such system helps to avoid having a biased classification score in the final step. For time-invariant functions, Supratak et al. [10] apply the residual connections and add temporal information extracted from the CNN to the sequential learning function by Bi LSTM. It is important to note that we also employ a residual CNN model in this thesis. One problem arising from training deep CNN networks is vanishing gradient. In its simplest form, a CNN consists of a convolution kernel applied to an input signal. A kernel consists of neural weights. These weights are to be learned and adjusted through training algorithm called backpropagation. A stacked representation of abstract features is obtained when various regions are activated by responding to convolution and pooling operations. Using the such procedure, an embedded representation of features is extracted from EEG signals. In this thesis, we represent the raw EEG signals as spectrogram images, and by doing so, we address the problem of sleep disorder recognition as an image classification problem.

3.3 Resnet Model

Whenever the numeral of layers in a DL network improves, computational complexity increases, and thus, accuracy saturates as the network converges toward the optimal solution. However, if the depth is too deep (too many layers), then the performance might degrade depending on the size of the data and the downstream task. A proper training and optimization strategy can improve the situation in such cases.

ResNets are a CNN-based architecture that benefits from skip connections and residual blocks to overcome the vanishing gradient problem. This makes it possible to have networks with as many as thousand convolution layers.

Residual blocks are the essential blocks of a ResNet network. To make very deep neural network architectures, intermediate inputs are added to the output of a group of convolution blocks using so-called skip connections. Alternatively, skip connections are known as identity mapping or residual connections. The following Figure illustrates a residual block, where X is the input to the Resnet block and can be an output from a previous layer, and $F(x)$ is a small network with few convolution layers.

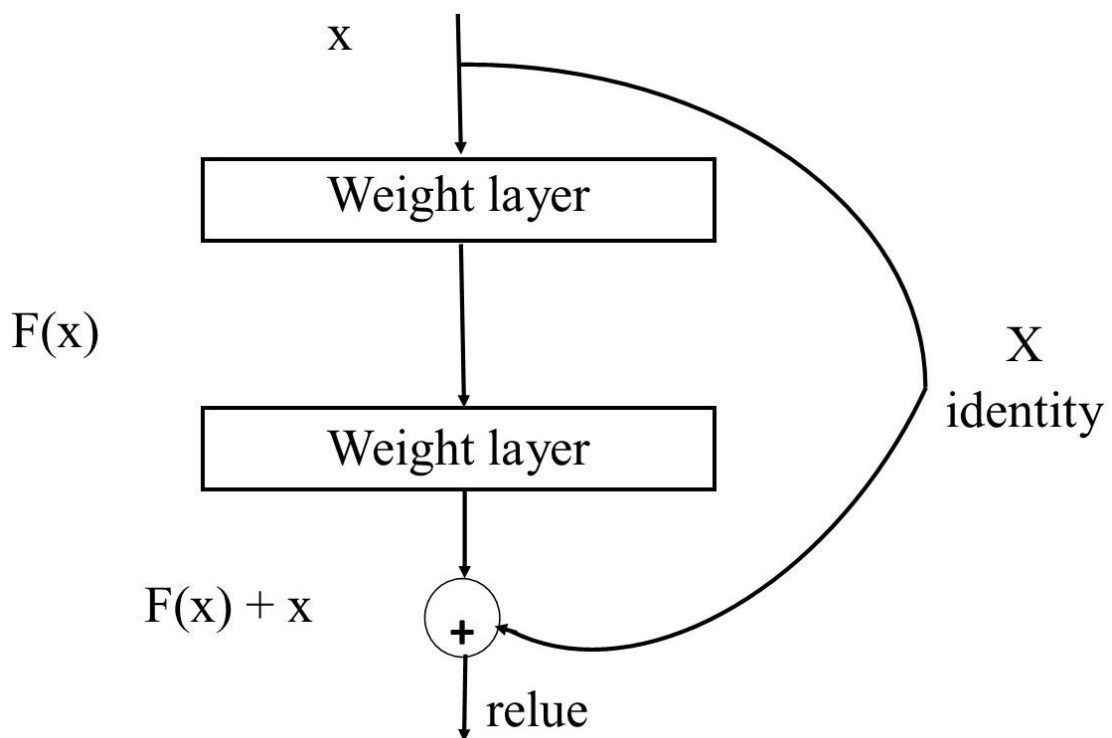


Figure 3.1: Resnet-Related Content [18]

3.4 Convolutional Architectures for EEG in the Literature

One known convolutional architecture that works explicitly on two-dimensional time-frequency representations of sleep data is proposed by [16]. In fact, the window size and layer structure of the architecture presented in this work is directly inspired by [16]. However, the input data for [16] is not spectrogram images. The spectrogram representation used in this research can be considered as the contribution of this work.

3.5 Experimental Methodology

This chapter describes the experimental methodology and details of the architecture used in this work. First, we explain the reason to employ selected architectures for this work. This is followed by a description of the datasets used. A specification of the convolutional architecture used to implement the classifier is then given. Finally, details of the training procedure are explained.

The architecture and all experiments described below are implemented in Matlab programming language.

3.6 Architecture Description

A fundamental innovation in this work is transforming time-series signals into a 2D image representation called spectrograms. In order to create a spectrogram, one should apply Fourier Transform (FT) on a subsegment of a signal. A spectrogram is a 2D image-based representation of the Short-Time Fourier Transform (STFT) of a signal where the horizontal axis introduces the time, and the vertical axis illustrates the signal's frequency. A second fundamental innovation can be considered as treating each separate signal as a channel of a color image. The time series of each EEG channel are independently converted to pseudocolor spectrogram images. A significant advantage of this technique is the ability to generate multichannel input data signals.

Such pipelines are extensively studied and optimized for image classification in machine-learning literature.

The spectrogram transformation above is convenient for representing the data in a low-level feature space. Spectrograms must be of sufficient resolution both in frequency and time domains. This way, fundamental frequency differences are detected as shifts along the frequency axis. Therefore the model assures some amount of frequency invariance in its performance; summarizing the above, the powerful technology of CNNs can be applied to spectrograms for the classification task. From an image processing perspective, the architecture is standard and is considered consistent with standard image processing literature.

3.7 Architecture Details

This section specifies spectrogram parameters and the parameter final result of this work.

3.7.1 Spectrogram

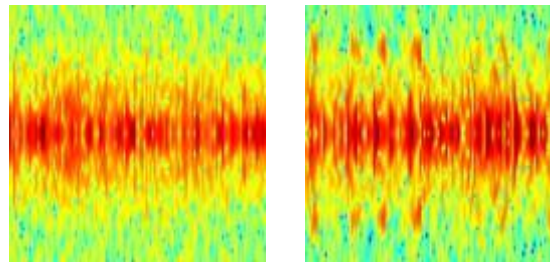
The parameter for generating spectrogram images was selected according to several criteria. First, The widths of the dominant spectral patterns should be approximately the same.

Second, there should be an overlap between successive time slices to smooth the spectral features. In order to maintain the details in pixel space, the signal is split into subsegments. The segmentation is shown to have significant improvement in classification results. Specifically, each 30-second chunk of the normalized EEG signal is sampled at 100 Hz according to the parameters in Table 3.1.

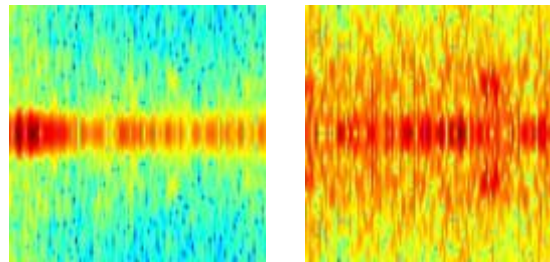
Table 3.1: Spectrogram parameters used.

Spectrogram	Fs	Freq.limitation	Time Resoulution
high freq	100Hz	[-20 20]	1S

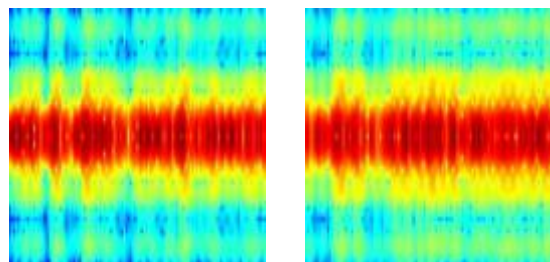
Spectrograms are normalized before being used as input for CNN architecture. Figure 3.2 shows examples of created spectrograms from various channels of an EEG signal.



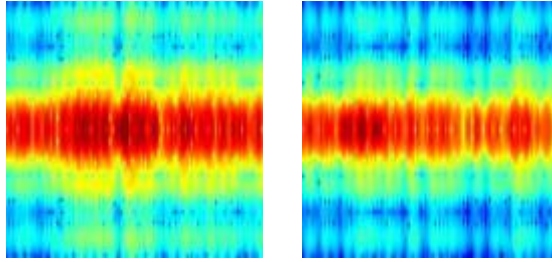
a) Mix signal(Fpz-Cz channel)



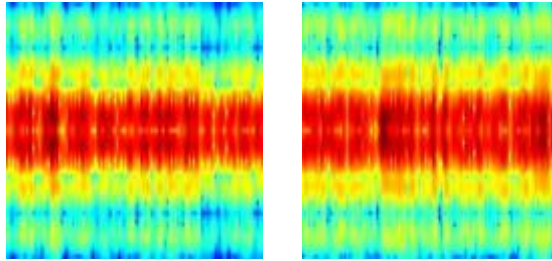
b) Mix Signal (Pz-Oz Channel)



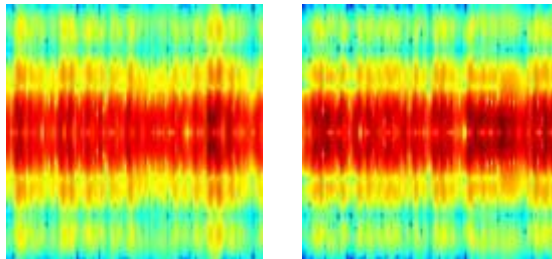
c) Delta signal (Fpz-Cz channel)



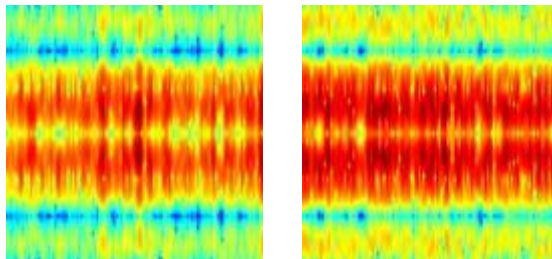
d) Delta signal (Pz-Oz channel)



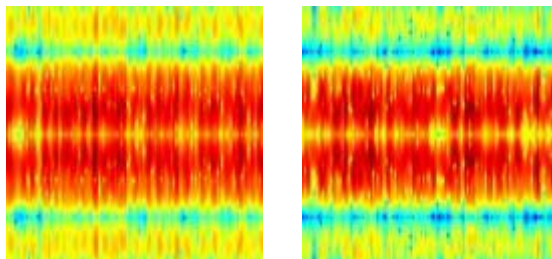
e) Theta signal (Fpz-Cz channel)



f) Theta signal (Pz-Oz channel)



g) Alpha signal (Fpz-Cz)



h) Alpha signal (Pz-Oz channel)

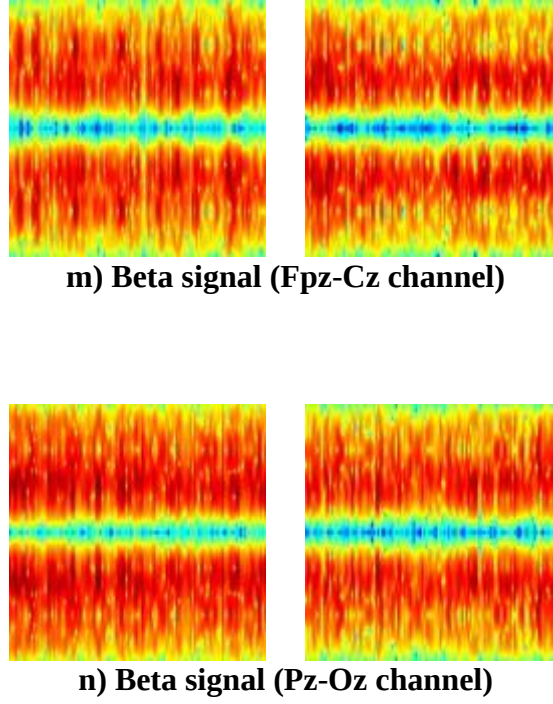


Figure 3.2: We can see some spectrogram images; In **part a** shows the Mix signal (Fpz-Cz channel) and two different stages, which is the R stage on (the left side) and S2 on (the right side). **Part b**, it shows the Mix signal (Pz-Oz channel) and two different stages, which is the S1 stage on (the left side) and W on (the right side); **part c** shows the S2 stage on (the left side) and S4 on (the right side) and also for **part d** shows S3 stage on (the left side) and R on (the right side); **part e** shows the S1 stage on (the left side) and S3 on (the right side) and also for **part f** shows R stage on (the left side) and W on (the right side); **part g** shows the S1 stage on (the left side) and S2 on (the right side) and also for **part h** shows S3 stage on (the left side) and R on (the right side); **part m** shows the S2 stage on (the left side) and S3 on (the right side) and also for **part n** shows S4 stage on (the left side) and W on (the right side).

3.7.2 Network Architecture

Input images are of size 128×128 . A remarkable difference with other available methods in the literature is the number of feature maps in the convolutional and fully connected layers. This means that the number of different spectral features in EEG is much smaller. Therefore the model complexity is proportionally lower, and consequently, the risk of overfitting is less than in other methods. Spectrograms are passed through separate convolutional layer stacks connected to a sequence of fully

connected layers with activations to generate a class label as the sleep stage of the corresponding subject.

3.8 Data Pre-Processing

Two EEG channels each were used for all recordings. Channels were Fpz-Cz, and Pz-Oz.

3.9 Training Methodology

The fundamental goals of this work are to train a classifier in this way. Works well for unseen patients. Therefore, training and validation are always strictly separated at the patient level. Report the final result. The next chapter contains the experimental results of validation and test data.

Generally speaking, 10% of the data is used as the test data set. And 90% percent for training and validation; the remaining data is sued for validation and training (90% for training and 10% for validation) is used. The dataset has 50 subjects, which resulted in 4200 spectrograms for six stages (W, R, S1-S4). From this number, 567 samples per category are used for training, and 63 images per category are used for validation and 70 samples are used for the test, which can be found in table 3.2.

Table 3.2: Number of images used for CNN

Stages/images	Total images	Test images	Training images	Validation images
R	700	70	567	63
S1	700	70	567	63
S2	700	70	567	63
S3	700	70	567	63
S4	700	70	567	63

W	700	70	567	63
---	-----	----	-----	----

The proposed model is shown in Figure 3.2

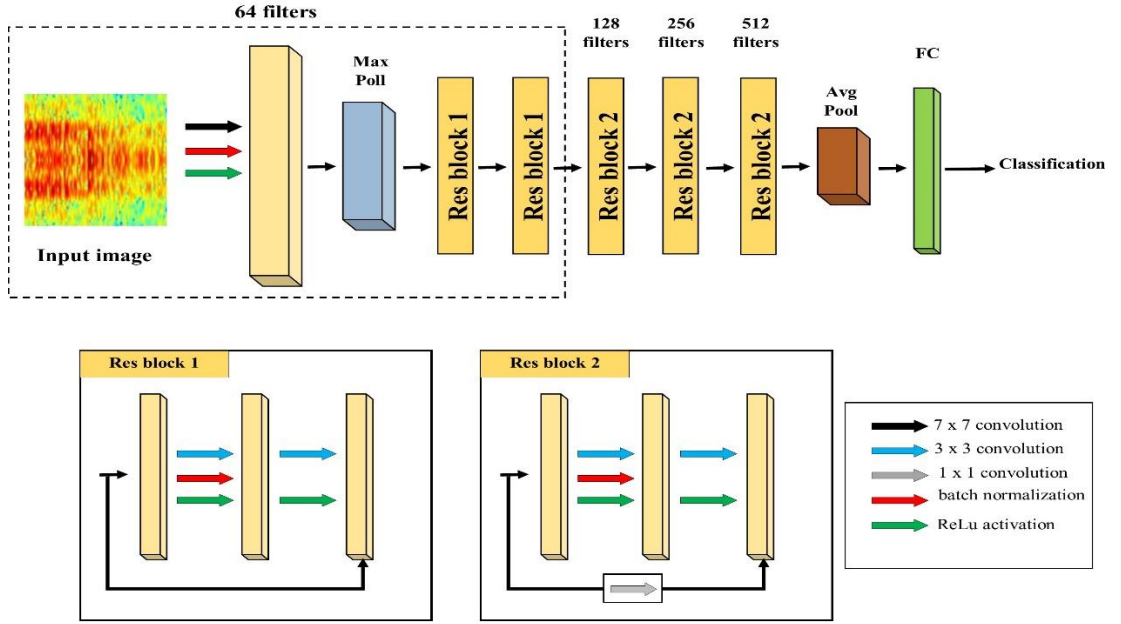


Figure 3.3: Resnet 18 block diagram

3.10 Expectations

Experts can easily categorize sleep stages by examining sleep stage recordings before and after a specific event. We expect that our proposed architecture competes with human expert decisions. In essence, instead of analyzing raw EEG signals, the network uses consecutive spectrograms to output a decision for the subject's sleep stage.

3.11 Fusion

Fusion is an approach by which one can improve the overall accuracy and performance of the corresponding classifier. As a result, we intend to combine or fuse the final confidence scores of classification obtained from various channels to ameliorate the

sleep stage classification. Correlating and merging information from multiple sources can usually provide more accurate conclusions than analyzing a single set of scores.

Chapter 4

RESULTS AND DISCUSSION

This chapter represents the conducted experiments and results of the developed architecture explained in the previous chapter. Various use cases are considered by using Mix signal channels and filter channels. Using such data input and Resnet 18 structure, we predict sleep stages.

4.1 Basic Performance

We computed four metrics as the classifier's output: F1-score, accuracy, precision, and recall. The last two are considered to be important when there is data imbalance. F1 score, recall, precision, and precision are defined mathematically for each output label. For a given output label, precision is the ratio of the number of true positive predictions over the total number of predictions for this label. The recall is the ratio of the number of true positive predictions to the number of positive predictions. These metrics are expressed as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.2 Performance

In addition to collecting recall, precision, accuracy, and F1-score parameters, we apply the score fusion method. We consider both channels, namely (Fpz-Cz and Pz-Oz). As for score fusion, we used fusion followed by an argmax of probabilities function.

More comprehensive information about the results is in Figure and Table 4.1.

Figure 4.1.1-4 shows mixed signal (Fpz-Cz and Pz-Oz) channels with training and loss graphs.

Figure 4.1.1 is train graphs for the Fpz-Cz channel. Used 70 images testing for each class, and the last number of training accuracy is 87.5%, validation accuracy is 70.11%, the number of epochs is 50, the learning rate is 0.005, and also for loss curves shown in figure 4.1.2 with validation loss is 1.11.

Figures 4.1.3 and 4.1.2 are train and loss curves for the Pz-Oz channel. All trains used the same number of testing (70 images for each class) with these parameters; the last number of training accuracy is 82.8, validation accuracy is 68.52%, number of epochs is 50, and the learning rate is 0.01, and for loss curves, we know validation loss is 0.8488.

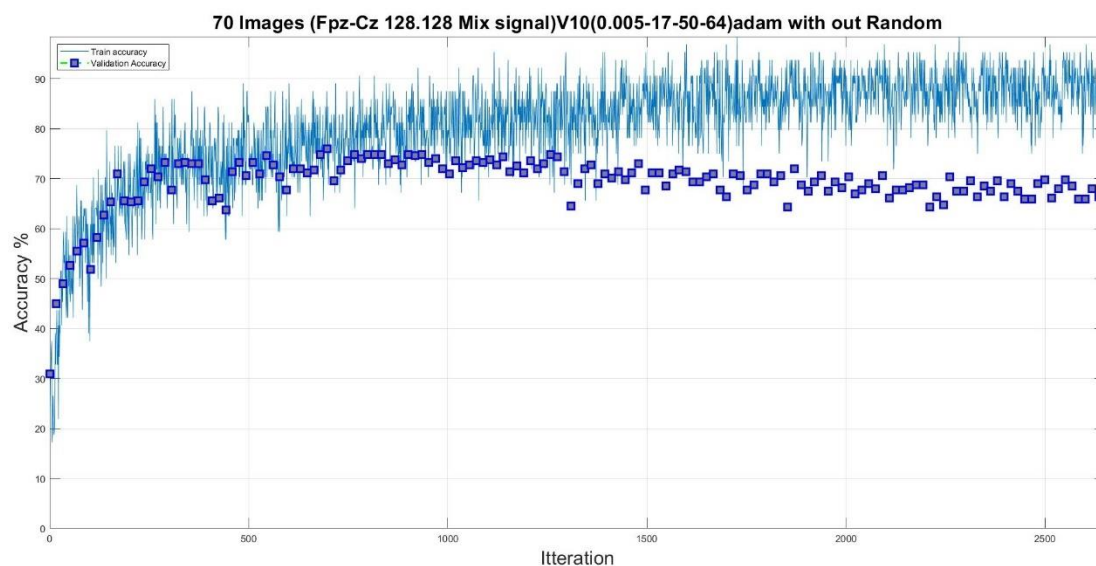


Figure 4.1: Training graphs for mix signal (Fpz-Cz channel)

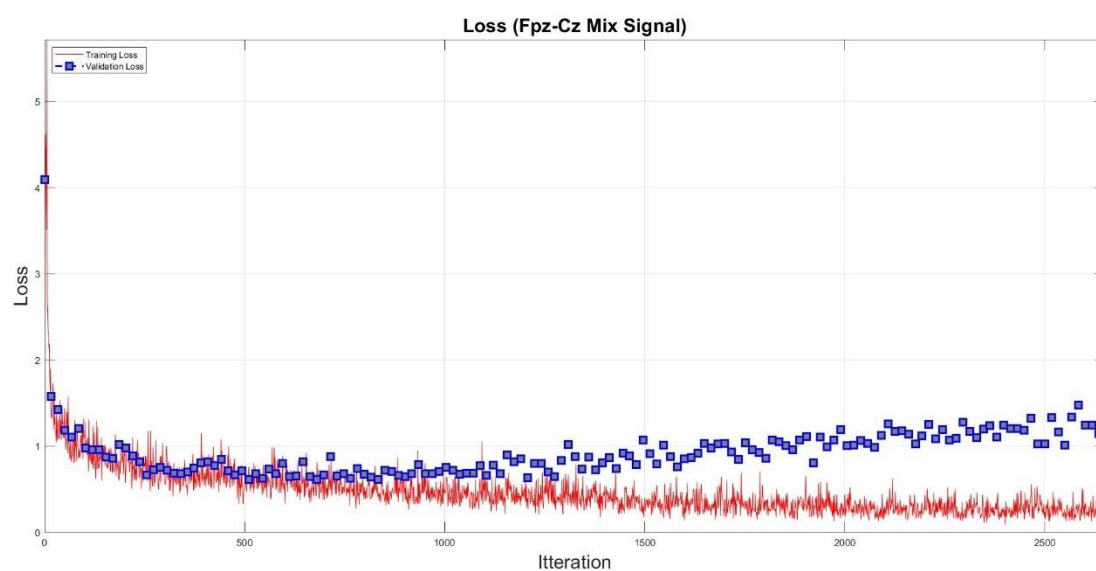


Figure 4.2: Loss graphs for mix signal (Fpz-Cz channel)

Table 4.1: Confusion matrix with the accuracy of Mix signal (Fpz-Cz channel)

Stages	R	S1	S2	S3	S4	W	Acc%
R	35	35					50%
S1	1	54	1	12	1	1	77.1%
S2	1	6	62	1			88.6%
S3		4		48	18		68.6%

S4		2	1	16	51		72.9%
W		1				69	98.6%

Table 4.2: Average precision, recall, accuracy, and F1-score for Mix signal (Fpz-Cz channel).

Precision	Recall	Accuracy	F1-score
70.79%	75.95%	75.95%	76.25%

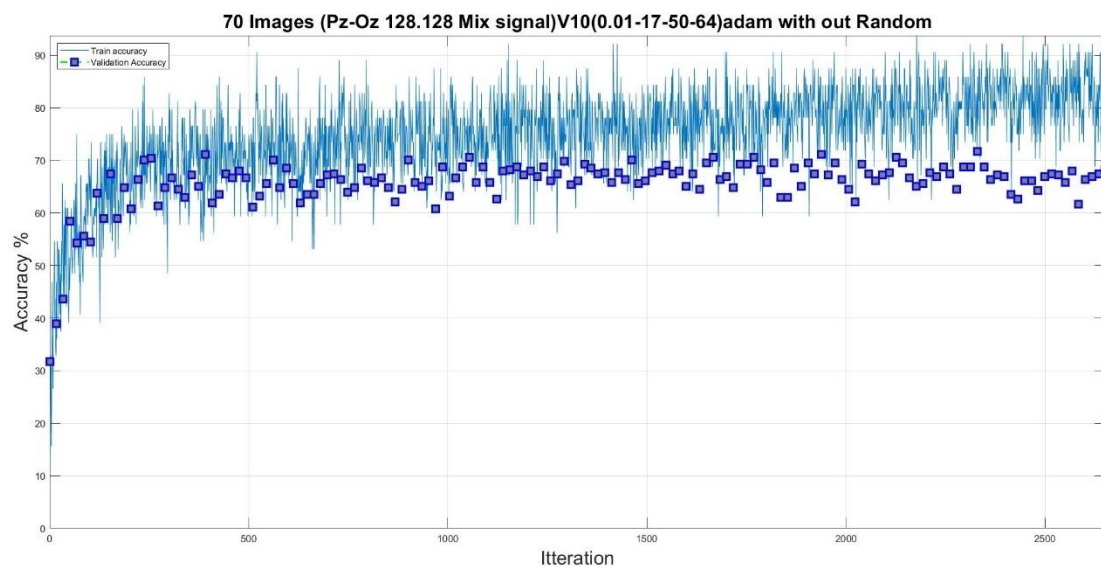


Figure 4.3: Training graphs for mix signal (Pz-Oz channel)

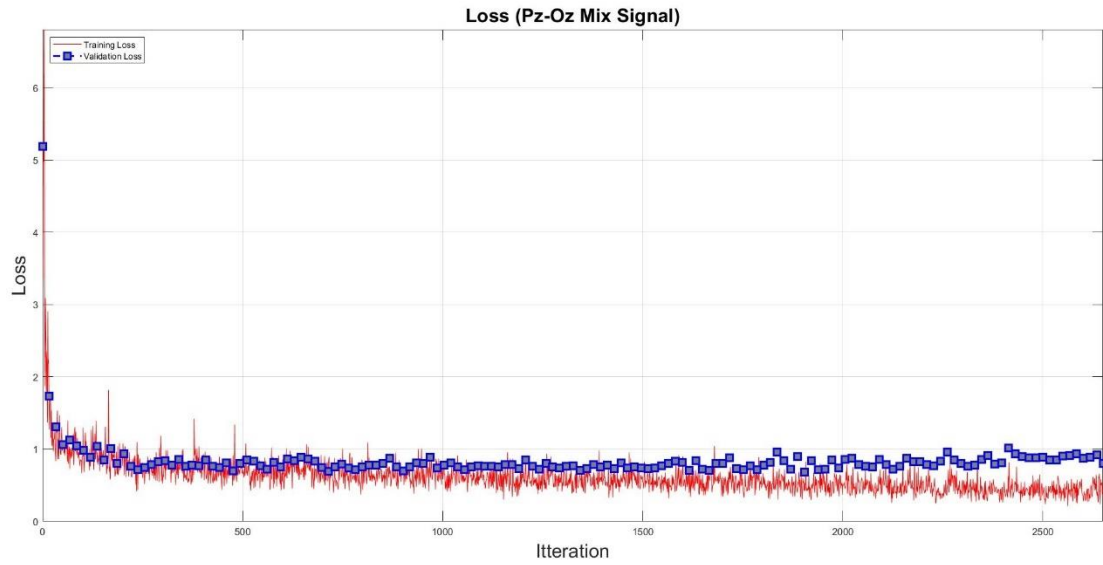


Figure 4.4: Loss Graphs for mix signal (Pz-Oz channel)

Table 4.3: Confusion matrix with accuracy of Mix signal (Pz-Oz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	47	23					67.1%
S1	3	35	10	13	6	3	50%
S2	2	4	63	1			90%
S3		2	1	34	32	1	48.6%
S4		4	2	20	44		62.9%
W		1				69	98.6%

Table 4.4: Average precision, recall, accuracy, and F1-score for Mix signal (Pz-Oz channel).

Precision	Recall	Accuracy	F1-score
70.36%	69.52%	69.52%	69.56%

Figure 4.1.5-8 shows Delta signal (Fpz-Cz and Pz-Oz) channels with training and loss graphs.

Figure 4.1.5 is train graphs for the Fpz-Cz channel(Delta signal). Used 70 images testing for each class, and the last number of training accuracy is 39.06%, validation accuracy is 41.79%, the number of epochs is 50, the learning rate is 0.01 , and also for loss curves shown in figure 4.1.6 with validation loss is 1.42.

Figures 4.1.7 and 4.1.8 are train and loss curves for the Pz-Oz channel(Delta signal). The last number of training accuracy is 45.3%, validation accuracy is 45.43%, number of epochs is 50, and the learning rate is 0.01, and for loss curves, we know validation loss is 1.44.

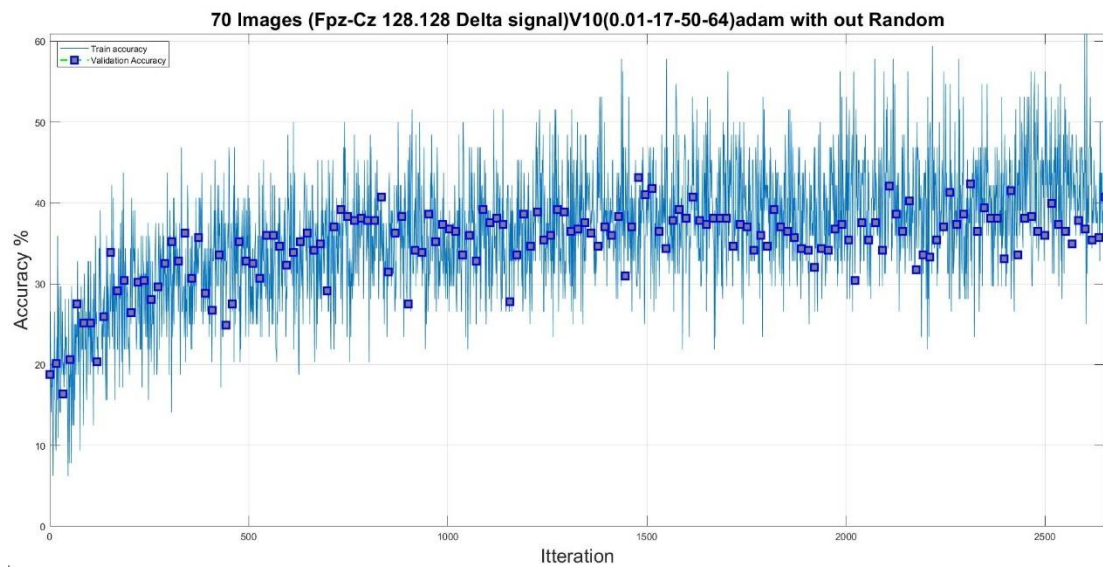


Figure 4.5: Training graph for delta signal (Fpz-Cz channel)

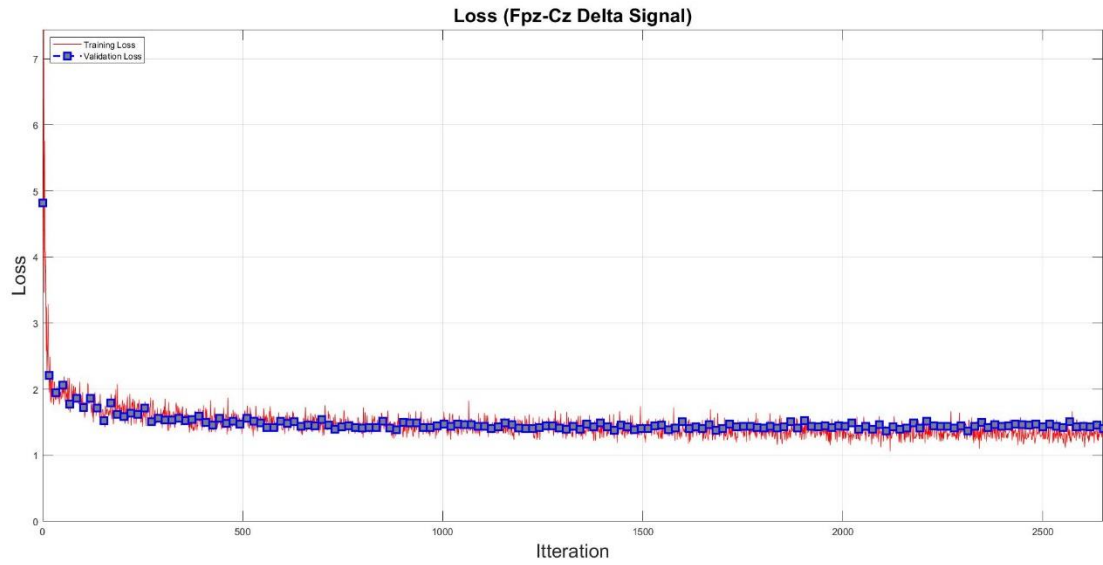


Figure 4.6: Loss graph for delta signal (Fpz-Cz channel)

Table 4.5: Confusion matrix with the accuracy of Delta signal (Fpz-Cz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	35	20	1	1		13	50%
S1	27	11	8	1	1	22	15.7%
S2	7		34	11	3	15	48.6%
S3	5		22	19	18	6	27.1%
S4			4	16	49	1	70%
W	5	3	15	8	6	33	47.1%

Table 4.6: Average precision, recall, accuracy and F1-score for Mix signal (Fpz-Cz channel).

Precision	Recall	Accuracy	F1-Score
41.89%	43.10%	43.10%	41.73%

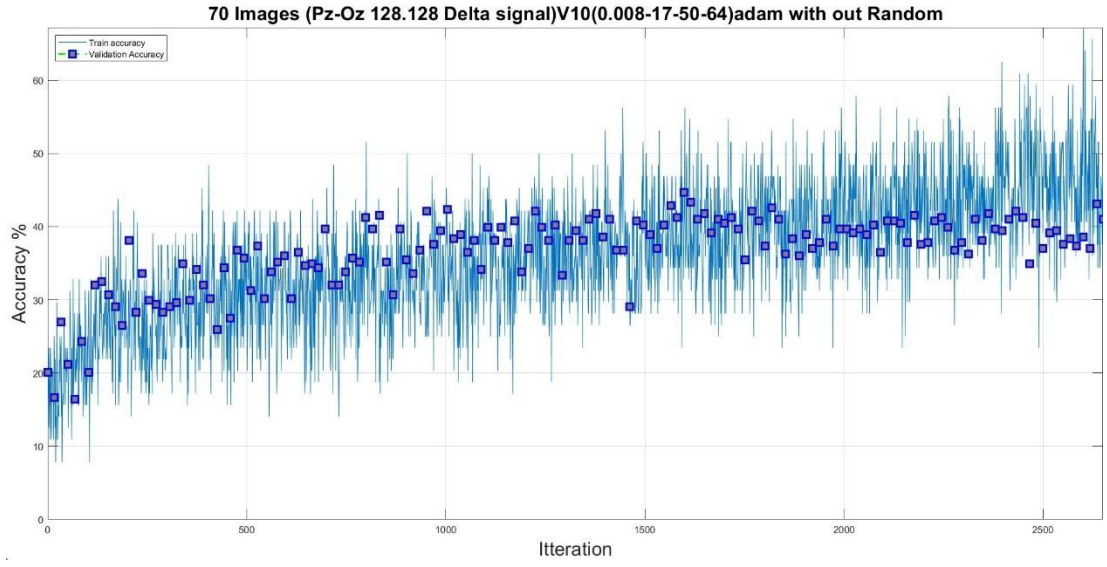


Figure 4.7: Training graph for delta signal (Pz-Oz channel)

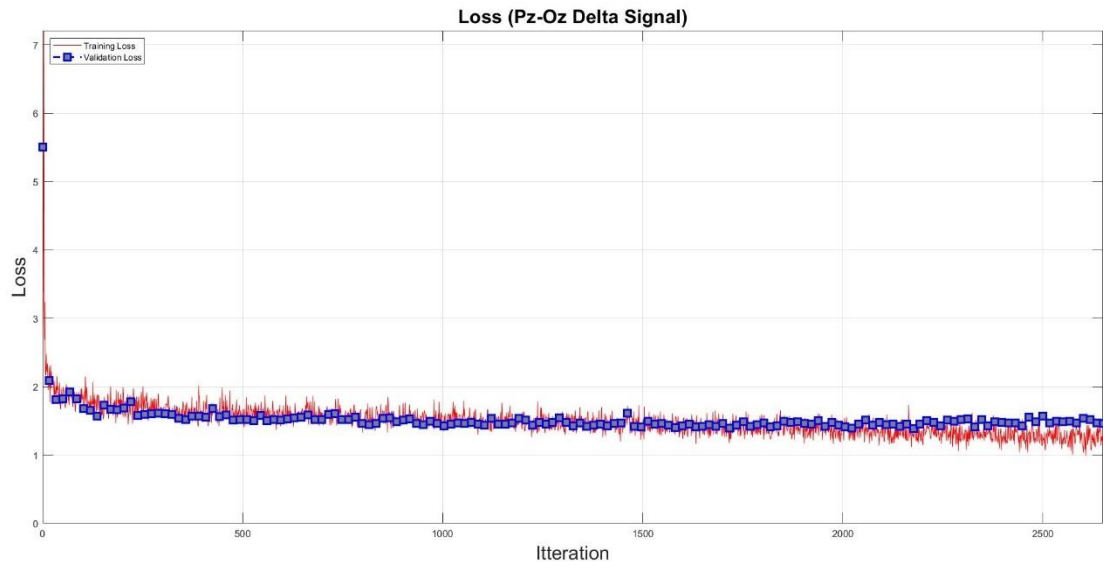


Figure 4.8: Loss graph for delta signal (Pz-Oz channel)

Table 4.7: Confusion matrix with accuracy of Delta signal (Pz-Oz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	19	6	7	1	3	34	27.1%
S1	21	35	3	2		9	50%
S2	17	8	18	11	11	5	25.7%

S3	14	10	14	13	15	4	18.6%
S4	5	4	8	13	38	2	54.3%
W	11	5	3	4	1	46	65.7%

Table 4.8: Average precision, recall, accuracy and F1-score for Delta signal (Pz-Oz channel).

Precision	Recall	Accuracy	F1-score
39.78%	40.24%	40.24%	39.37%

Figure 4.1.9-12 shows Theta signal (Fpz-Cz and Pz-Oz) channels with training and loss graphs.

Figure 4.1.9 is train graphs for the Fpz-Cz channel(Theta signal). Used 70 images testing for each class, and the last number of training accuracy is 7187%, validation accuracy is 35.71%, the number of epochs is 50, the learning rate is 0.008 , and also for loss curves shown in figure 4.1.10 with validation loss is 2.04.

Figures 4.1.11 and 4.1.12 are train and loss curves for the Pz-Oz channel (Theta signal). The last number of training accuracy is 85.93%, validation accuracy is 25.92%, number of epochs is 50, and the learning rate is 0.001, and for loss curves, we know validation loss is 3.42.

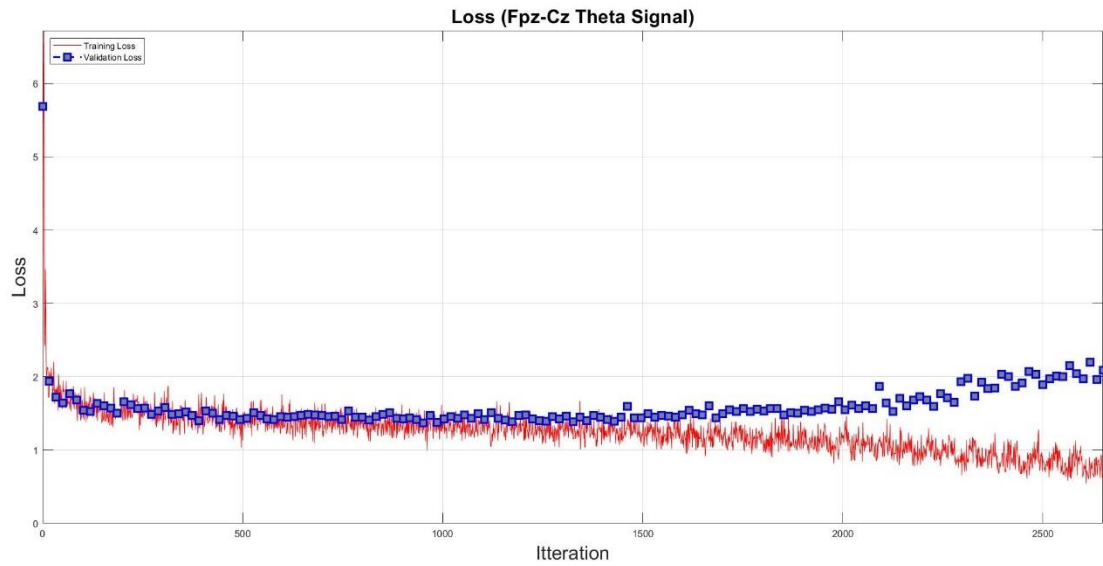
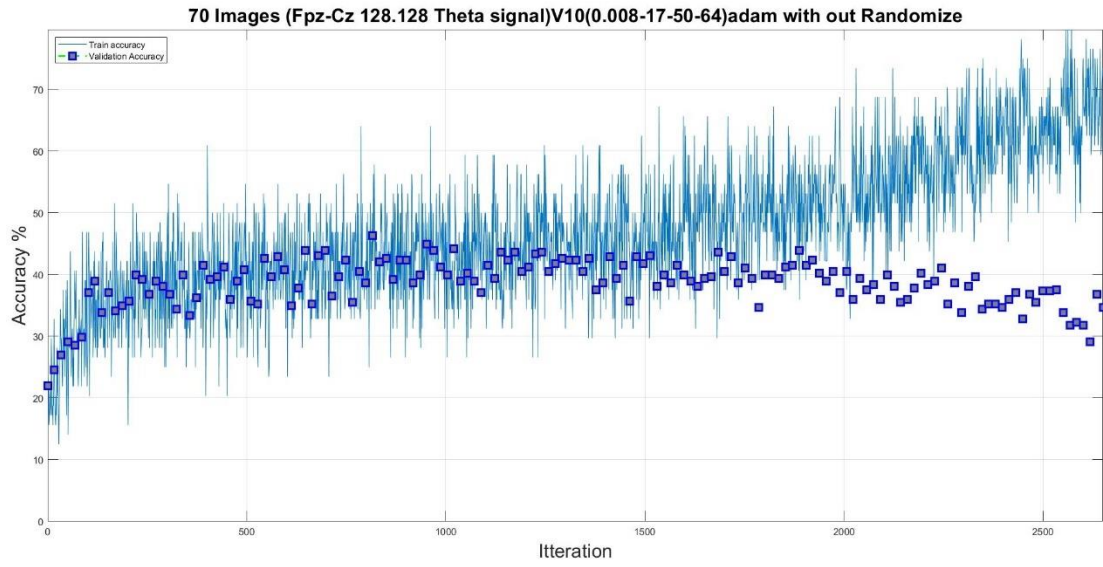


Table 4.9: Confusion matrix with accuracy of Theta signal (Fpz-Cz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	20	19	11	5	3	12	28.6%
S1	15	16	20	8	4	7	22.9%
S2	14	10	19	8	7	12	27.1%
S3	2	9	8	7	37	7	10%

S4	2	9	6	11	41	1	58.6%
W	2	4	2	2	5	55	78.6%

Table 4.10: Average precision, recall, accuracy and F1-score for Tehta signal (Fpz-Cz channel).

Precision	Recall	Accuracy	F1-score
38.36%	38.81%	38.81%	37.97%

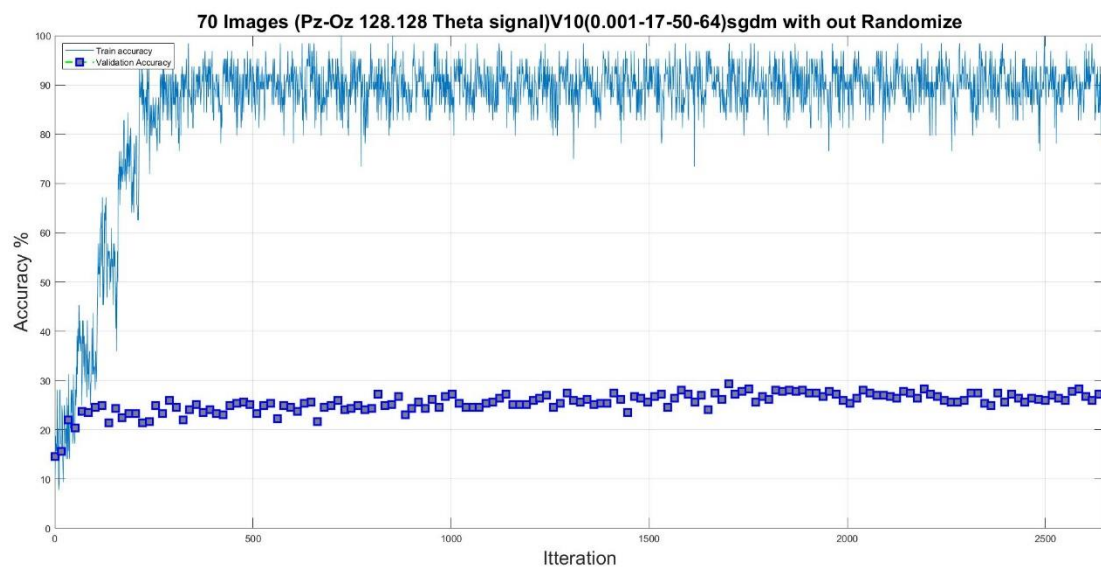


Figure 4.11: Training graph for theta signal (Pz-Oz channel)

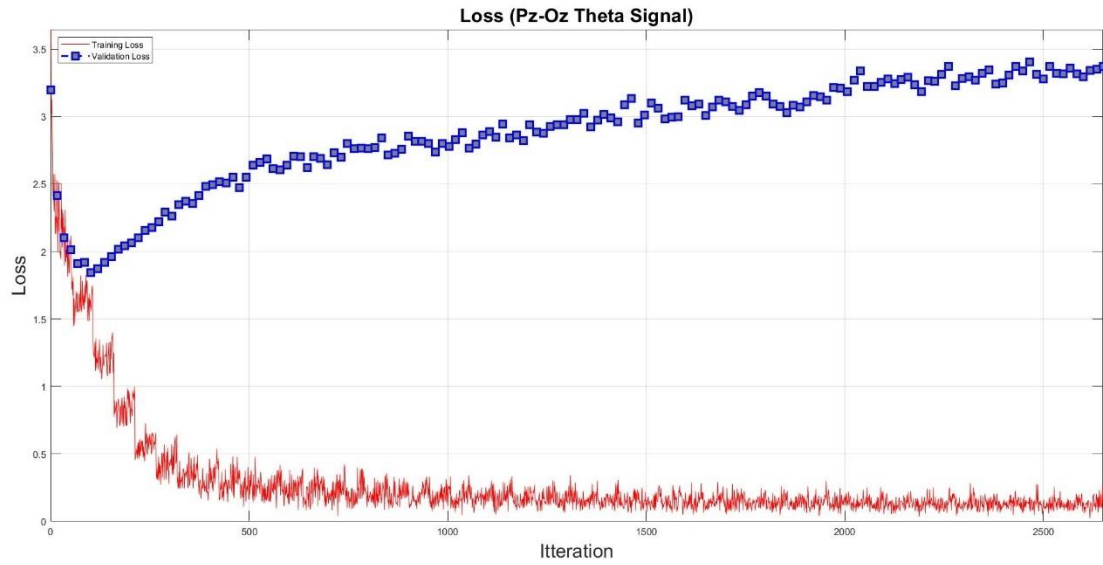


Figure 4.12: Loss graph for theta signal (Pz-Oz channel)

Table 4.11: Confusion matrix with accuracy of Theta signal (Pz-Oz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	20	25	10	6	4	5	28.6%
S1	10	22	16	5	11	6	31.4%
S2	9	10	21	7	10	13	30%
S3	7	10	9	12	30	2	17.1%
S4	4	5	16	14	23	8	32.9%
W	10	6	16	11	11	16	22.9%

Table 4.12: Average precision, recall, accuracy and F1-score for Theta signal (Pz-Oz channel).

Precision	Recall	Accuracy	F1-score
27.51%	27.14%	27.14%	26.98%

Figure 4.13-16 shows Alpha signals (Fpz-Cz and Pz-Oz) channels with training and loss graphs.

Figure 4.1.13 is train graphs for the Fpz-Cz channel (Alpha signal). Used 70 images testing for each class, and the last number of training accuracy is 57.81%, validation accuracy is 45.50%, the number of epochs is 50, the learning rate is 0.01 , and also for loss curves shown in figure 4.1.14 with validation loss is 1.47.

Figures 4.1.15 and 4.1.16 are train and loss curves for the Pz-Oz channel (Alpha signal). The last number of training accuracy is 26.56%, validation accuracy is 34.12%, number of epochs is 50, and the learning rate is 0.01, and for loss curves, we know validation loss is 1.69.

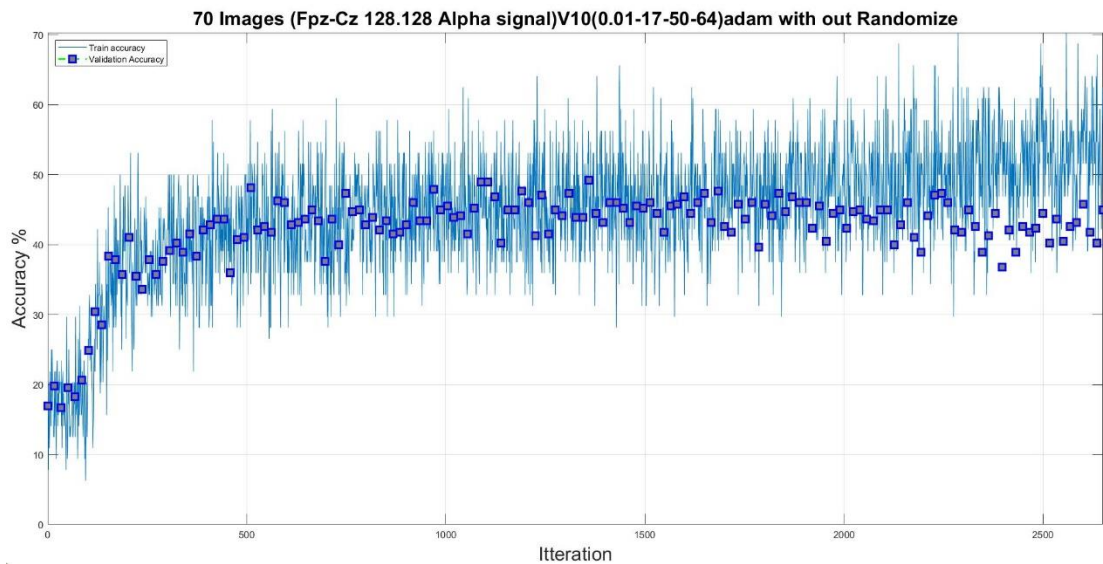


Figure 4.13: Training graph for alpha signal (Fpz-Cz channel)

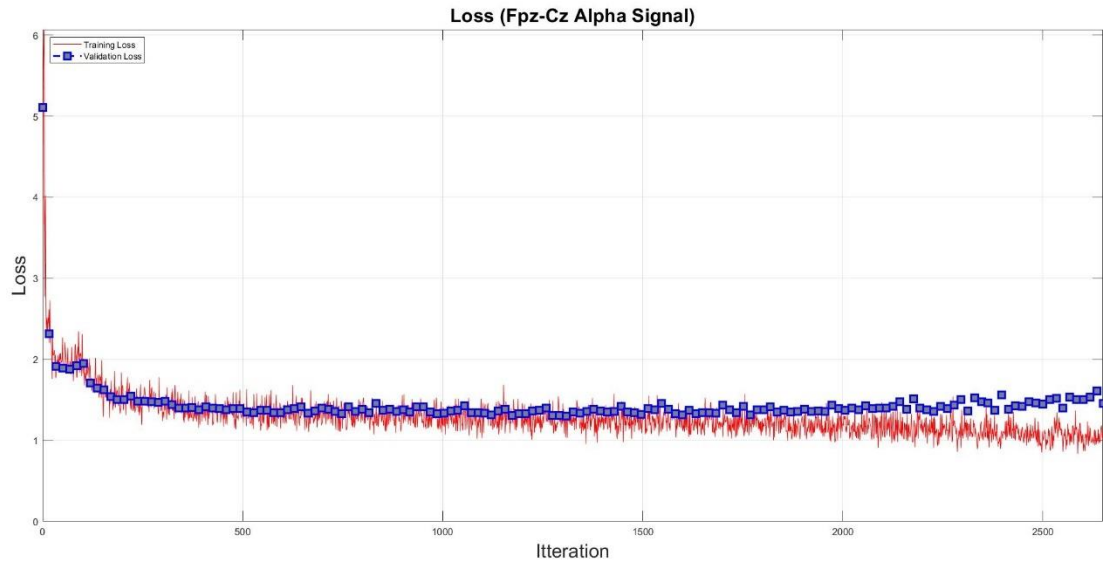


Figure 4.14: Loss graph for alpha signal (Fpz-Cz channel)

Table 4.13: Confusion matrix with accuracy of Alpha signal (Fpz-Cz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	39	3	13	1	2	12	55.7%
S1	6	14	32	3	4	11	20%
S2	15	8	23	3	8	13	32.9%
S3	5	3	25	11	22	4	15.7%
S4	3	2	25	5	34	1	48.6%
W	1	1	4	1		63	90

Table 4.14: Average precision, recall, accuracy and F1-score for Alpha signal (Fpz-Cz channel).

Precision	Recall	Accuracy	F1-score
45.92%	43.81%	43.81%	42.03%

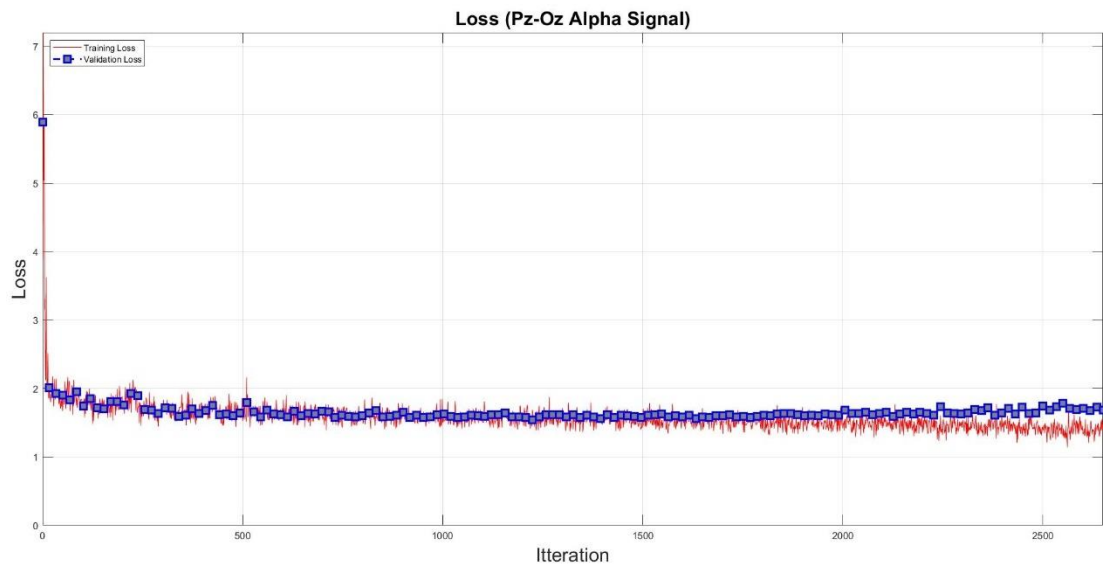
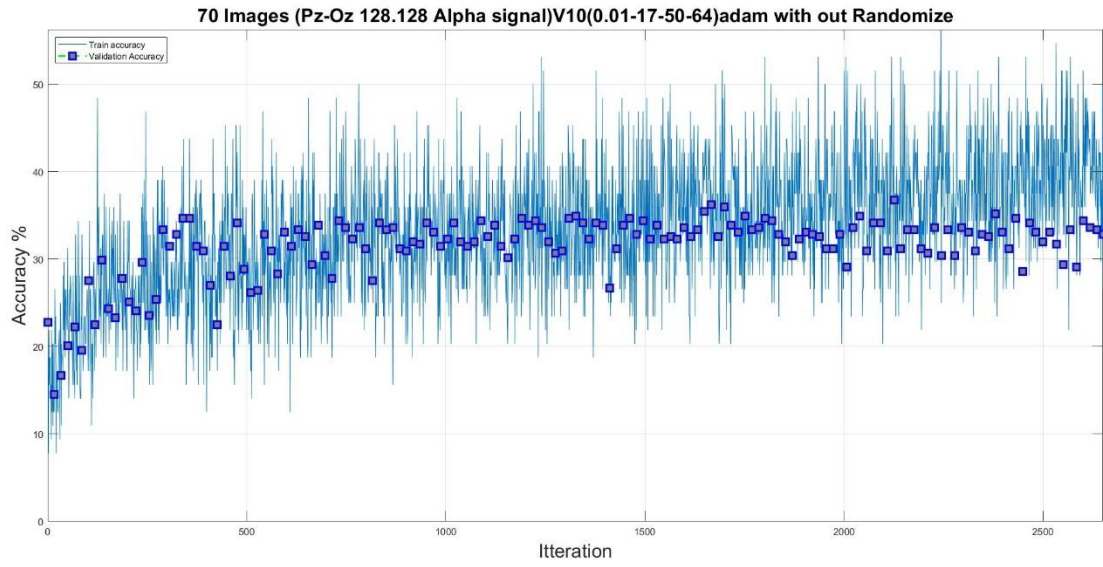


Table 4.15: Confusion matrix with accuracy of Alpha signal (Pz-Oz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	2	20	14	4	7	23	2.9%
S1	2	24	9	8	24	3	34.3%
S2	7	10	18	5	4	26	25.7%

S3	2	8	9	5	45	1	7.1%
S4		1	14	15	35	5	50%
W	24	9	3	4		30	42.9%

Table 4.16: Average precision, recall, accuracy and F1-score for Alpha signal (Pz-Oz channel).

Precision	Recall	Accuracy	F1-score
23.72%	27.14%	27.14%	24.77%

Figure 4.1.17-20 shows Beta signals (Fpz-Cz and Pz-Oz) channels with training and loss graphs.

Figure 4.1.17 is train graph for the Fpz-Cz channel (Beta signal). Used 70 images testing for each class, and the last number of training accuracy is 86%, validation accuracy is 41.53%, the number of epochs is 50, the learning rate is 0.001 , and also for loss curves shown in figure 4.1.18 with validation loss is 2.42.

Figures 4.1.19 and 4.1.20 are train and loss curves for the Pz-Oz channel (Beta signal). The last number of training accuracy is 49.20%, validation accuracy is 47.35%, number of epochs is 50, and the learning rate is 0.01, and for loss curves, we know validation loss is 1.70.

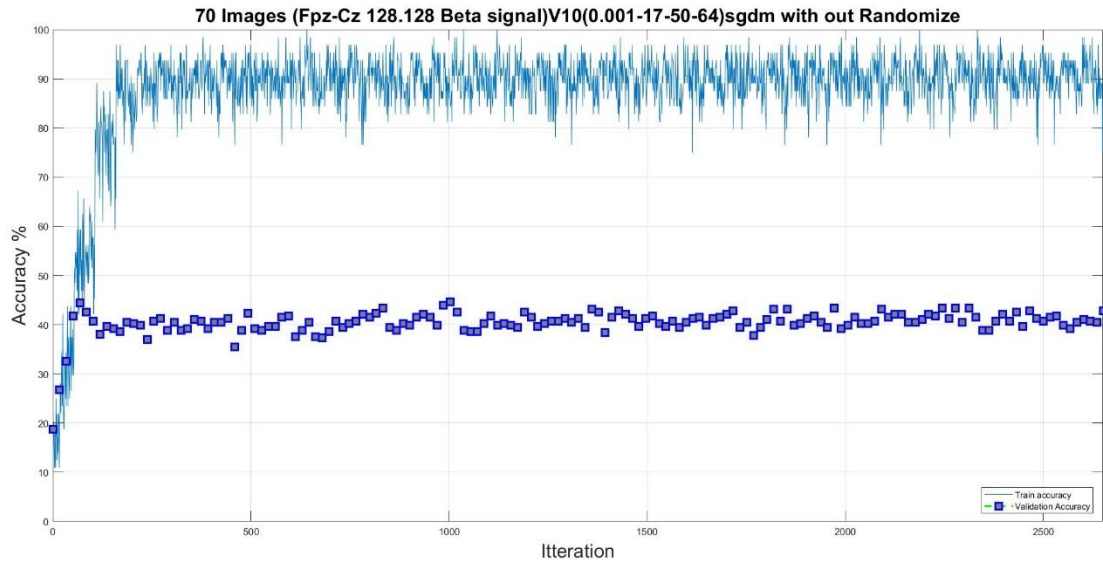


Figure 4.17: Training graph for beta signal (Fpz-Cz channel)

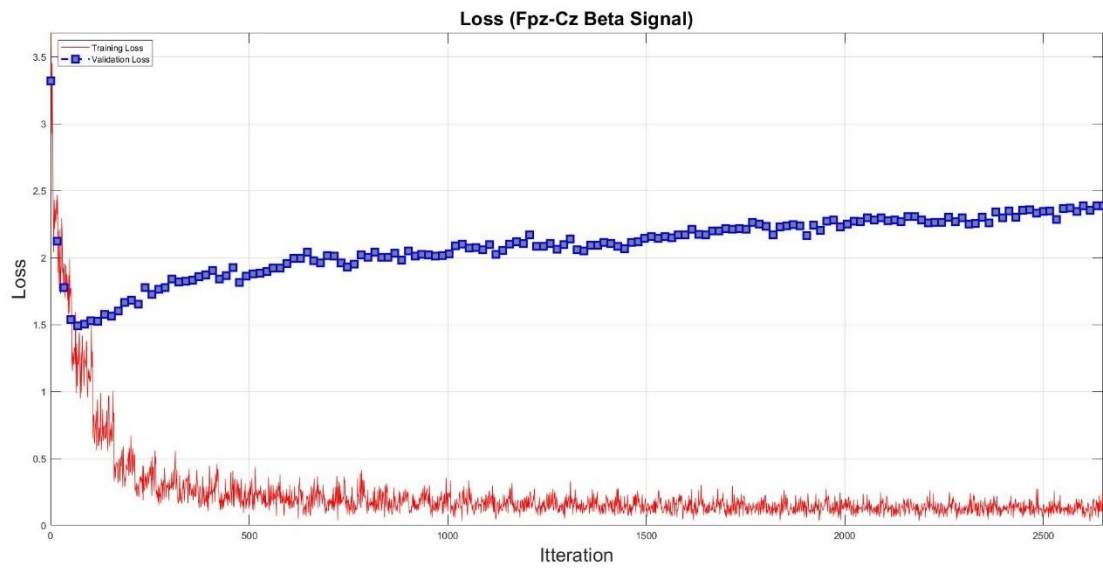


Figure 4.18: Loss graph for beta signal (Fpz-Cz channel)

Table 4.17: Confusion matrix with accuracy of Beta signal (Fpz-Cz channel).

Stage	R	S1	S2	S3	S4	W	Acc%
R	25	22	4	8	4	7	35.7%
S1	14	19	18	8	4	7	27.1%
S2	2	6	38	7	16	1	54.3%
S3	11	8	5	29	13	4	41.4%

S4	5	7	5	15	25	13	35.7%
W	10	4		6	4	46	65.7%

Table 4.18: Average precision, recall, accuracy and F1-score for Beta signal (Fpz-Cz channel).

Presicion	Recall	Accuracy	F1-score
42.83%	43.33%	43.33%	43.03%

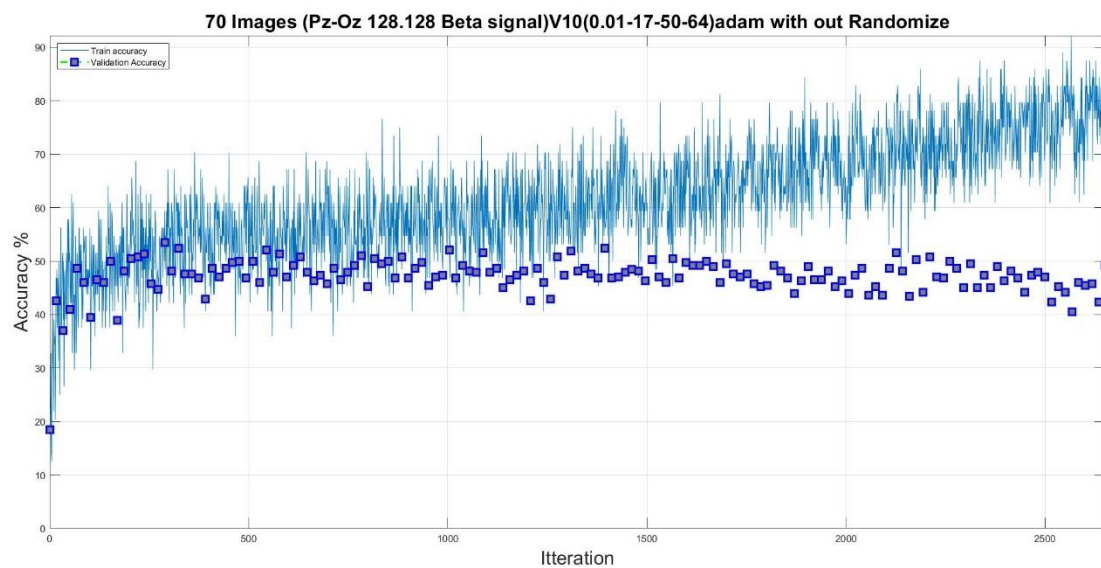


Figure 4.19: Training graph for beta signal (Pz-Oz channel)

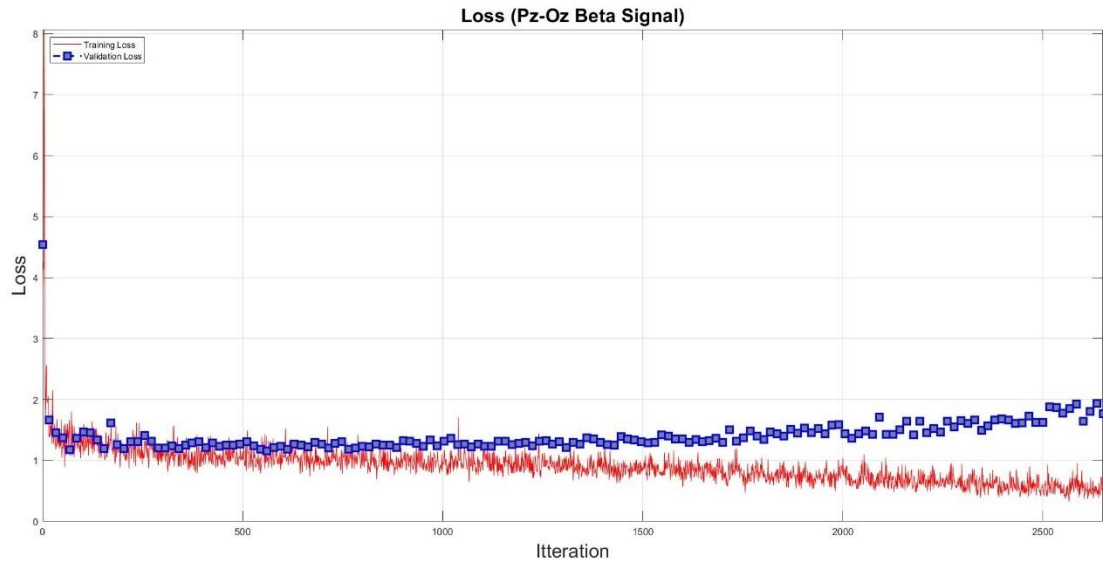


Figure 4.20: Loss graph for beta signal (Pz-Oz channel)

Table 4.19: Confusion matrix with accuracy of Beta signal (Pz-Oz channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	23	34	3	1	3	6	32.9%
S1	12	36	12	1	2	7	51.4%
S2	5	6	29	9	21		41.4%
S3	17	7	5	19	21	1	27.1%
S4	8	14	18	5	25		35.7%
W	9	21	4			36	51.4%

Table 4.20: Average precision, recall, accuracy and F1-score for Beta signal (Pz-Oz channel).

Precision	Recall	Accuracy	F1-score
43.91%	40%	40%	40.46%

4.3 Score Fusion Performance

Results for score fusion approach are reported. The idea was to add scores of two channels and apply the argmax function to predict the class label.

Table 4.21: Confusion matrix with accuracy of Mix signal (combined channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	42	28					60%
S1	2	53		13	1	1	75.7%
S2	2	5	63				90%
S3		3		45	22		64.3%
S4		4	1	15	50		71.4%
W						70	100%

Table 4.22: Average precision, recall, accuracy and F1-score for Mix signal (Combined channel).

Precision	Recall	Accuracy	F1 score
79.24%	76.90%	76.90%	77.27%

Table 4.23: Confusion matrix with accuracy of Delta signal (combined channel).

Srages	R	S1	S2	S3	S4	W	Acc%
R	29	10	2		1	28	41.4%
S1	22	31	8		1	8	44.3%
S2	13	3	33	11	6	4	47.1%
S3	10	2	19	16	20	3	22.9%
S4	1		2	12	54	1	77.1%

W	8	4	6	3	4	45	64.3%
----------	---	---	---	---	---	----	-------

Table 4.24: Average precision, recall, accuracy and F1-score for Delta signal (Combined channel).

Precision	Recall	Accuracy	F1 score
42.26%	49.52%	49.52%	48.52%

Table 4.25: Confusion matrix with accuracy of Theta signal (combined channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	32	32		2	1	3	45.7%
S1	1	50	4	10	3	2	71.4%
S2	1	7	55	1	3	3	78.6%
S3	4	5	2	37	21	1	52.9%
S4	1	3	4	13	48	1	68.6%
W				1		69	98.6%

Table 4.26: These values are for average precision, recall, accuracy and F1-score for Delta signal (Combined channel).

Precision	Recall	Accuracy	F1 score
71.09%	69.29%	69.29%	68.95%

Table 4.27: Confusion matrix with accuracy of Alpha signal (combined channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	36	9	12		2	11	51.4%

S1	4	22	21	8	9	6	31.4%
S2	12	8	25	3	5	17	35.7%
S3	1	5	16	8	38	2	11.4%
S4	1	2	22	5	38	2	54.3%
W	2	2	2	1		63	90%

Table 4.28: Average precision, recall, accuracy and F1-score for Alpha signal (Combined channel).

Precision	Recall	Accuracy	F1 score
45.22%	45.71%	45.71%	43.61%

Table 4.29: Confusion matrix with accuracy of Beta signal (combined channel).

Stages	R	S1	S2	S3	S4	W	Acc%
R	32	26	3	3	3	3	45.7%
S1	17	29	13	1	3	7	41.4%
S2	2	5	36	8	19		51.4%
S3	13	9	5	27	14	2	38.6%
S4	9	10	12	11	21	7	30%
W	7	9		1	1	52	74.3%

Table 4.30: Average precision, recall, accuracy and F1-score for Beta signal (Combined channel).

Precision	Recall	Accuracy	F1 score
47.62%	46.90%	46.90%	46.94%

4.4 Final Concolusion

These results are for 6 classses (R, W, S1-4). In table 4.3, it can be seen that the accuracy percentage of score fusion was higher than the single channel method. In Channel 1 column and channel two column wich means accuracies and for After fusion added up probabilities and take the argmax.

Table 4.31: Total accuracy for each signal (for 6 classes)

Signal/Chanell	Channel 1	Channel 2	After Fusion
<i>Mix sgnal</i>	75.95%	69.52%	76.00%
<i>Delta</i>	43.09%	40.23%	49.00%
<i>Theta</i>	38.8%	27.14%	68.00%
<i>Alpha</i>	43.8%	29.00%	45.00%
<i>Beta</i>	43.33%	40.00%	47.00%

4.5 Comparise performance

In this part, we will see how our method compares with other similar studies shown in Table 4.4.

In the comparison table, most of them with five stages are classified, but in this study we try to used six stages of comparison. Usually in sleep stage classification study, it can be combine S3 and S4 together, but for deeper research we separated S3 and S4 from each other. Also the accuracy for five stages after fusion (sum ruls) is 86%.

Table 4.32: Performance Comparison of the proposed and alternative methods

<i>STUDY</i>	<i>SUBJECTS</i>	<i>INPUT SIGNALS</i>	<i>DL MODEL</i>	<i>W(%)</i>	<i>S1(%)</i>	<i>S2(%)</i>	<i>S3(%)</i>	<i>S4(%)</i>	<i>R(%)</i>	<i>OVERALL ACC (%)</i>
JADVAH et al [29]	42	Fpz-Cz	Pre-trained CNN	94.50	36.00	92.70	79.90	-	86.90	84.70
SUPRATAK et al [30]	20	Fpz-Cz	CNN+RNN	83.40	50.10	81.70	94.20	-	83.90	82.00
MOUSAVI et al [31]	20	Fpz-Cz	CNN+RNN	90.60	54.50	82.70	88.90	-	88.70	84.30
PROPOSED	50	Fpz-Cz	CNN	98.60	74.30	85.70	91.40	-	63.00	82.57
PROPOSED	50	Pz-Oz	CNN	97.10	64.30	88.60	94.30	-	61.40	81.14
PROPOSED	50	Fpz-Cz	CNN	98.60	77.10	88.60	68.60	73.00	50.00	76.00
PROPOSED	50	Pz-Oz	CNN	98.60	50.00	90.00	48.60	63.00	67.10	69.50

Chapter 5

CONCLUSIONS

5.1 Summary of Contributions

This study presented a deep learning approach for automated sleep stage classification using convolutional neural network and spectrograms that are created from multichannel EEG signals. The CNN architecture was trained and validated on dataset of approximately 50 EEG recordings from patients collected in a hospital setting. It is important to note that we achieved a competitive performance without extracting hand-crafted features in a separate pre-processing step.

The experimental results show that results achieved by automated classifiers can perform close to that human experts recognizing sleep disorder in clinics by practice.

5.2 Observations on Classifier Performance

Overall, the performance of the accuracy developed in this work is appropriate for screening, and its application in this field is a reasonable goal. However, drawbacks must be addressed before considering such automated approaches for practical usage. These drawbacks include but are not limited to having poor performance on classification of S3 and S4 stages. One solution to address this might be considering the two stage as a single class label [2]. Also, the performance of such classifiers degrades when using single channel data.

5.3 Future Directions

In this study we attempted to show that CNN-based image processing techniques can be used for sleep stage classification once the EEG signals are converted to time-frequency domain images (spectrograms). Future direction of this research line includes applying one dimensional CNNs to raw EEG signals and fusion of results with the current 2D classification method. We believe applying 1D CNNs to EEG signals can automatically extract features that are salient and robust with respect to time domain and they can be an appropriate replacement for hand-crafted features.

More importantly, this method should be extended to be compatible with non-EEG recordings. Indeed, one of the strengths of the deep method is the ability to integrate information from different sources. Thus, the combination of other commonly seen signals, such as EMG, ECG, or other time series recording, should be attempted. In general, for clinical practice, developing a software package or tool with a pre-trained architecture is the most reliable approach.

Alternative minor directions comprise working on "drop channels" of the EEG signals where a subset of EEG channels is used in each sample. This makes the model capable of learning from various channels available in different clinical deployment sites.

In general, comparing with other available neural architectures like Xception and Inception networks and fusion with complex neural networks may bring more improvement to this field.

5.4 Methodological Improvements

No matter how good, the architecture will not be used in practice if clinicians lack confidence in its results. The method of preventing bagging has been developed for

this reason. For the architecture to be truly viable. The architecture needs to be trained on a large enough set of patients from different hospitals, whose measurements are made by different technicians with different tools. If the program's performance on a completely new data set is quantified, it is reasonable to believe it can act as a reliable or assistant tool.

REFERENCES

- [1] E. A. Wolpert, “A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects,” *Archives of General Psychiatry*, vol. 20, no. 2, pp. 246–247, 1969.
- [2] Jr.Lee-Chiong,et al., “Sleep Medicine: Essentials and Review”, Oxford University Press, pp.41-43, 2008
- [3] Medscape website <https://emedicine.medscape.com/article/1188142-overview>
- [4] 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.
- [5] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, “Automatic sleep stage scoring with single-channel EEG using convolutional neural networks,” *ArXiv*, vol. abs/1610.01683, 2016.
- [6] O. Yildirim, U. B. Baloglu, and U. R. Acharya, “A deep learning model for automated sleep stages classification using PSG signals,” *International journal of environmental research and public health*, vol. 16, no. 4, p. 599, 2019.
- [7] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, “Understanding neural networks through deep visualization,” *arXiv preprint arXiv:1506.06579*, 2015.

- [8] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, “A convolutional neural network for sleep stage scoring from raw single-channel EEG,” *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.
- [9] J. Wang, Y. Zhang, Q. Ma, H. Huang, and X. Hong, “Deep learning for single-channel EEG signals sleep stage scoring based on frequency domain representation,” in *HIS*, 2019.
- [10] A. Supratak, H. Dong, C. Wu, and Y. Guo, “Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel EEG,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.
- [11] A. I. Humayun, A. S. Sushmit, T. Hasan, and M. I. H. Bhuiyan, “End-to-end sleep staging with raw single-channel EEG using deep residual convnets,” *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–5, 2019.
- [12] J. Friedman, T. Hastie, and R. Tibshirani, “The elements of statistical learning”, volume 1. Springer series in statistics Springer, Berlin, 2001.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning”, MIT Press, 2016.
- [14] M. Arif Wani ,Farooq Ahmad Bhat , Saduf Afzal , Asif Iqbal Khan, “Advances in Deep Learning”, Spring, 2020

- [15] <https://vitalflux.com/cnn-basic-architecture-for-classification-segmentation>
- [16] P.Bashivan, I.Rish, M.Yeasin, and N.Codella, “Learning representations from eeg with deep recurrent-convolutional neural networks”. arXiv preprint arXiv:1511.06448, 2015.
- [17] C. Iber, et al, “The AASM manual for the scoring of sleep and associated events: rules”, terminology and technical specifications. American Academy of Sleep Medicine, 2007.
- [18] <https://neurohive.io/en/popular-networks/resnet>
- [19] K.He, X.Zhang, S.Ren, and J.Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.”, In Proceedings of the IEEE international conference on computer vision, pages 1026-1034, 2015.
- [20] G.NaveenSundar , D.Narmadha ,A.AmirAnton Jone , K.MartinSagayam, Hien Dang, Marc Pomplun “Automated sleep stage classification in sleep apnoea using convolution neural network”, Volume 26, 2021, 100724
- [21] H.ElMoaqet, M.Eid, M.Ryalat, T.Penzel, “A Deep Transfer Learning Frameworkfor Sleep Stage Classification with Single-Channel EEG Signals ”, Sensors 2022, 22(22), 8826

- [22] K. Fukushima, "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193–202, 1980.
- [23] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology*, vol. 195, no. 1, pp. 215–243, 1968.
- [24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [26] E. Oropesa, H. Cycon, and M. Jobert, "Sleep stage classification using wavelet transform and neural network." *International computer science institute* (1999)
- [27] R. Khan, "Why Using Mean Squared Error(MSE) Cost Function for Binary Classification is a Bad Idea?" 11. Nov 2019. url: <https://towardsdatascience.com/why-using-mean-squared-error-mse-cost-function-for-binaryclassification-is-a-bad-idea-933089e90df7>.

- [28] N.Liu, et al. “Learning a convolutional neural network for sleep stage classification.” Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on. IEEE, 2017
- [29] P.Jadhav, G.Rajguru, D. Datta, S. Mukhopadhyay, “Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network”, Biocybern. Biomed. Eng. 2020, 40, 494–504.
- [30] A.Supratak, H.Dong, C.Wu, Y.Guo, “DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Trans.”, Neural Syst. Rehabil. Eng. 2017, 25, 1998–2008.
- [31] S.Mousavi, F.Afghah, U.R Acharya, “SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach.”, PLoS ONE 2019, 14, e0216456