

Enhanced Accuracy for Motor Imagery Detection Using Deep Learning for BCI



Author

Ayesha Sarwar

00000205843

Supervisor

Dr. Kashif Javed

DEPARTMENT OF ROBOTICS AND INTELLIGENT MACHINE
ENGINEERING
SCHOOL OF MECHANICAL & MANUFACTURING
ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD
NOVEMBER,2020

Enhanced Accuracy for Motor Imagery Detection Using Deep Learning for BCI

Author

AYESHA SARWAR

00000205843

A thesis submitted in partial fulfillment of the requirements for the degree of
MS Robotics and Intelligent Machine Engineering

Thesis Supervisor:

Dr. Kashif Javed

Thesis Supervisor's Signature: _____

**DEPARTMENT OF ROBOTICS AND INTELLIGENT MACHINE
ENGINEERING
SCHOOL OF MECHANICAL & MANUFACTURING
ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD
NOVEMBER, 2020**

Thesis Acceptance Certificate

It is certified that the final copy of MS Thesis written by Ayesha Sarwar (Registration No. 00000205843), of SMME (School of Mechanical & Manufacturing Engineering) has been vetted by undersigned, found complete in all respects as per NUST statutes / regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil Degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature: _____

Name of Supervisor: Dr. Kashif Javed

Date: _____

Signature (HOD): _____

Date: _____

Signature (Principal): _____

Date: _____


National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **Ms. Ayesha Sarwar**, NUST Regn # **205843**, Titled: “**Enhanced Accuracy for Motor Imagery Detection using Deep Learning for BCI**” be accepted in partial fulfillment of the requirements for the award of **MS Robotics and Intelligent Machines Engineering degree**. (Grade____)

Examination Committee Members

1. Name: **Dr. Hasan Sajid** Signature: _____

2. Name: **Dr. Umar Ansari** Signature: _____ 

3. Name: **Dr. Karam Daad Kallu** Signature: _____ 

Supervisor's name: **Dr. Kashif Javed** Signature: _____ 

Co-Supervisor's name: **Dr. Muhammad Jawad Khan** Signature: _____

Head of Department

Date

COUNTERSIGNED

Date: _____

Principal

FORM TH-4

Declaration

I certify that this research work titled “*Enhanced Accuracy of Motor Imagery Detection using Deep Learning for BCI*” is my work. The work has not been presented anywhere for the assessment. The material that has been used from other resources has been properly acknowledged/referred.

Signature of Student

AYESHA SARWAR

MS RIME

00000205843

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for plagiarism. Turnip report endorsed by Supervisor is attached.

Signature of Student

AYESHA SARWAR

00000205843

Signature of Supervisor

Dr. Kashif Javed

Copyright Statement

- Copyright in the text of this thesis rests with the student author. Copies (by any process) either in full or if extracts, may be made only by instructions given by the author and lodged in the Library of NUST School of Mechanical & Manufacturing Engineering (SMME). Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights which may be described in this thesis is vested in NUST School of Mechanical & Manufacturing Engineering (SMME), subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of SMME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST School of Mechanical & Manufacturing Engineering (SMME), Islamabad.

Acknowledgments

I am grateful to the creator of the whole universe, the Great Allah Almighty, who directed me in every field of life, at each step for enlightening my mind to do my work properly. Indeed, without your priceless support and advice, I could have done nothing. All the people, my parents, friends, or any other person who supported me throughout my course study were because of Allah's will.

I am deeply grateful to my dear parents who brought me up when I was unable to speak or walk. They continue supporting me at every step in every field of my life.

Finally, I am very thankful to my supervisor Dr. Kashif Javed and my co-supervisor Dr. Muhammad Jawad Khan, and GEC members Dr. Omer Ansari, Dr. Hassan Sajid, and Dr. Karam Daad Kallu for their help throughout my work.



Dedicated to my great parents, supervisor, and co-supervisor, whose immense support and coordination have led me to this incredible achievement.

ABSTRACT

The Brain-Computer Interface (BCI) is a system where the human brain and a hardware device interact instantly. This transfers the brain data recorded directly to the computer that can be used for external system control. There are four key components of the BCI method, namely the acquisition of signals, preprocessing of acquired signals, extraction, and classification of features. In conventional machine learning algorithms, the accuracy achieved is negligible and not up to the mark to classify motor imagery data for multiple classes. The main explanation for this is that features are manually selected, and we are unable to get certain features that result in greater precision. For classifying the multi-class motor imagery (MI) data, we have implemented deep learning algorithms in this work. Two different approaches have been explored in this study: Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM). We evaluate the accuracy of classification on two datasets, i.e. Competition for BCI III, dataset IIIa and competition for BCI IV, dataset IIa. The results showed that deep learning algorithms provide higher accuracy outcomes than conventional machine learning algorithms. LSTM outperforms the ANN and the deep learning classifier gives 96.2 percent average classification accuracy.

Key Words:

Brain-computer interface; motor imagery; artificial neural network; long-short term memory, classification.

Table of Contents

THESIS ACCEPTENCE CERTIFICATE	i
FORM- TH4	ii
DECLARATION	ii
PLAGARISM CERTIFICATE (Turnitin Report)	iii
COPYRIGHT STATEMENT	iv
ACKNOWLEDGEMENTS	v
ABSTRACT	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES	xi
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION	1
Motivation.....	2
Objective.....	3
CHAPTER 2: LITERATURE REVIEW	5
CHAPTER 3: METHODOLOGY	9
Experimental Studies.....	10
BCI Competition III, Dataset IIIa.....	10
Paradigm	10
Preprocessing.....	13
BCI Competition IV, Dataset IIa.....	16
Paradigm.....	16
Preprocessing.....	18
Adopted Methodology.....	19
Deep Learning Method.....	20
Artificial Neural Network (ANN).....	21
Long Short-Term Memory (LSTM).....	22
CHAPTER 4: RESULTS	24
CHAPTER 5: CONCLUSION & DISCUSSIONS	32
REFERENCES	33
ORIGINALITY REPORT	35

List of Figures

Figure 1: Basic components of the BCI system.....	2
Figure 2: The time scheme for the paradigm	10
Figure 3(a): Raw signal part 1.....	11
Figure 3(b): Raw signal part 2.....	12
Figure 3(c): Raw signal part 3.....	12
Figure 4(a): Single class trial part 1.....	14
Figure 4(b): Single class trial part 2.....	14
Figure 4(c): Single class trial part 2.....	15
Figure 5(a): Raw signal.....	15
Figure 5(b): Filtered Signal.....	16
Figure 6: The time scheme for the paradigm.....	16
Figure 7(a): Raw signal of one subject's data.....	17
Figure 7(b): Single class trial.....	17
Figure 8: Adopted methodology flow chart.....	19
Figure 9: Summary of LSTM.....	22
Figure 10(a): Accuracy graph.....	24
Figure 10(b): Loss graph.....	25
Figure 11(a): Accuracy graph.....	26
Figure 11(b): Loss graph.....	26
Figure 12(a): Accuracy graph.....	28
Figure 12(b): Loss graph.....	28
Figure 13(a): Accuracy graph.....	29
Figure 13(b): Loss graph.....	30

List of Tables

Table 1: Nomenclature.....	13
Table 2: Hyper-parameters for ANN.....	20
Table 3: Hyper-parameters for LSTM.....	22
Table 4: Classification accuracies for dataset IIIa with different window sizes.....	27
Table 5: Classification accuracies for dataset IIa with different window sizes.....	30
Table 6: Comparison of the different machine learning algorithms with deep learning algorithms on dataset IIIa, BCI competition III and dataset IIa, BCI competition IV.....	34

INTRODUCTION

Motivation

Brain-Computer Interface (BCI) is an evolving exploration area that expects to enhance the standard and level of computer-based applications on humans. Brain-Computer Interface offers ways to communicate your feelings without any vocal contact with the external system. For the last two, three decades BCI has gained researchers' attention and a lot of work started in this field. In 1973 French neurophysiologist Jaques Vidal first made use of the word Brain-Computer Interface while predicting the possibility of combining the brain activity and the power of the computer.

A brain-computer interface-based computer system is a system in which movements related to muscles are replaced with device commands that are produced during the translation of brain signals. These types of systems are of great importance when used for locked-in state (LIS) patients that are suffering from movement related disorders and are not able to communicate with the outside world [1]-[3]. Other than the healthcare field, there are applications of BCI systems like gaming, virtual reality, every-day human-computer interaction (HCI), and performance monitoring for humans [4] – [6].

A Brain-Computer Interface system comprises of the components: A BCI signal acquisition, signal processing, extraction of features, and classification [7] as shown in figure 1. In the first section, the electrical activity of the brain is gathered by performing some voluntary tasks. Brain signals may be obtained using two different methods: invasive and non-invasive. For signal acquisition purposes, the electrode is mounted in the brain's scalp invasive method. In a non-invasive technique, signals are acquired without conducting any surgical intervention.

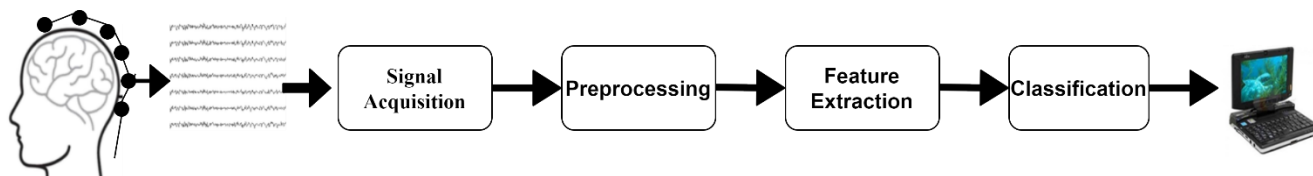


Figure 1: Basic Components of BCI System

Electroencephalogram (EEG) is one of the most commonly used signal acquisition methods due to its non-invasive nature, low cost, and ease of usage. German psychiatrist Hans Berger in 1924 for the first-time recorded EEG signals. EEG is a technique in which the brain's electrical signals are measured by placing an array of electrodes on the human brain's scalp. The brain signals gather through EEG are then pre-processed. The next step after the pre-processing of signals is the feature extraction. Then the features vector is given as input to train a classifier. As a result, the classifier classifies the electrical signal according to the user's motor imagery movement.

Objective

The prime objective of this research is to develop a novel algorithm for the motor imagery data for better classification of signals in the field of BCI. Different types of challenges are faced during BCI studies like experiment design, signals acquisition features extraction, and training of data. All of these steps require time. Since the non-invasive methods of data acquisition are more prone to noise and artifacts like muscles movement and eye blink becomes a challenge in the classification of such data. Furthermore, the desired feature that is best for the classification of these data signals is not known. In traditional machine learning approaches the best features are collected by hit and trial method where you first evaluate the features and after evaluation, the best ones are chosen for classification purpose.

In this study, we are optimizing the classification part of the BCI system by using different deep learning methods. In the previous studies, machine learning methods have provided better classification accuracies[9],[10], and [11]. The major issue is the extraction of features[8]. Moreover, the problem of features extraction in BCI is also resolved in deep learning techniques because in deep learning networks features can be extracted from the raw data. In this work, we are going to implement Artificial Neural Networks (ANN) and Recurrent Neural Network (RNN). There are different algorithms in RNN but we implemented Long Short-Term Memory (LSTM) architecture. The study will be split into the following major objectives to achieve the goal of this research:

- Literature Review

This review will discuss the already existing Machine Learning approaches for the classification of EEG signals for this research. Testing methods and other requirements will also be researched and discussed in this review.

- Dataset and Methodology:

In methodology, there will be a discussion about the data set used in this research and the architectures i.e. Artificial neural network (ANN) and Recurrent Neural Network (LSTM) that are used to classify those data sets.

- Results:

In the result section, there will be the results of both the architectures used in this research

- Conclusion:

In the end, there will be conclusions regarding this research and future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Understanding the EEG, ANN, and LSTM

In this study two different algorithms: Artificial Neural Network and Long-Short Term Memory are used for the classification of EEG datasets in four classes. Therefore, it is mandatory to understand EEG signals, working of ANN and LSTM. In this chapter, we describe the details of EEG, ANN, and LSTM.

2.1.1 Electroencephalography (EEG)

For capturing the brain's activity there are different ways [12]: electrooculography (EOG), electroencephalography (EEG), near-infrared spectroscopy (NIRS), and function near-infrared spectroscopy (fNIRS) that is the most recent one. EEG is a mostly used method for imaging that can be used for Brain-Computer Interfaces (BCI) because of its non-invasive nature, ease of use, portability, and cheap. It is less sensitive to artifacts, which make it ideal for long-term studies. It also makes it possible for dysfunctional subjects such as children or partially or completely locked-in state patients. In the detection of hemodynamic brain response, spatial resolution is higher than the detection of neural stimulus-response in EEG [13],[14]. There is the following type of noises present in an EEG signal.

(i) Motion Artifacts

Different motion artifacts occur in the form of peaks and changes in baseline in data the major reason is due to muscle movements that can occur during the recording of signals and due to improper scalp attachment, they should be removed because if they are too many, the entire data is rejected. In situations where they cannot be ignored, for example, where the dataset is limited, small, or where recordings from subjects can not be prevented. The best solution is to remove those artifacts and restore the signal. Different approaches are used to remove motion artifacts, like recording additional data on the subject's movement using referenced channels.[13],[14]and[15].

(ii) Instrumental Noise

Using basic low-pass filtering methods, instrumental noise is a random noise that can be removed. After data conversion to the frequency domain, a method such as Moving Average and cutting off higher frequencies are used. The sensitivity of these methods needs to be manually determined to avoid data distortion [13],[16].

(iii) Physiological Noise

When coping with noise sources for physiological use, potentials of skin's outer layer, and ionic features. It is important to regard the capacity of sweat glands also. The solution to handle this problem is to reduce the skin potential and improve the signal to noise ratio, by using abrasive cream on the skin.[11],[17] and[18].

Power line noise can be removed using a notch filter. For removing the noise from the raw dataset bandpass filter can be used. The lower cut-off frequency used in this dataset is 0.5 Hz and a higher cut-off frequency at 100 Hz.

2.1.2 Artificial Neural Network (ANN)

ANN is a network of feed-forward neurons comprising various layers [22]. To update the weights, backpropagation is used. In ANN architecture, there are various types of layers: an input layer, rectified linear units (ReLU) layer, dense layers which are also called fully connected layers. The input layer of the network is then transformed into a one-dimensional array and connected to eight hidden fully connected dense layers. In each hidden layer, the number of a neuron depends on input in every layer. Every input and output is linked via learnable weights in these layers. Non-linear activation function such as ReLU is also passed through each connected layer. Usually, in the final flatten layer there are the same amount of nodes as the number of classes or groups to which the input data must be classified. For the last layer, the activation function is chosen very carefully and

generally differ from previous fully connected layers. In this work the activation function used for the last layer is SoftMax.

2.1.3 Recurrent Neural Network (LSTM)

Long Short-Term Memory, one of the Recurrent Neural Network types, is the other architecture we use to explore the classification of EEG signals. Initially, the Recurrent Neural Network was represented by the study group of Schmidhuber and won eight international competitions in pattern recognition and machine learning [25],[26]. LSTM network provides the best results for data classification of time series and LSTM provides better results compared to traditional methods according to previous studies [27][28]. It operates on the concept of saving a layer's output and feeding that output to the next input layer, which assists in predicting the outcome of the layer. LSTM is one of the RNN types capable of learning from observation, which is an advantage of LSTM over other types of RNNs and neural networks. The key thing about LSTM is the cell state that can be modified, i.e. it can be removed or added. In LSTM, there are three gates: the forget gate, the input gate, and the output gate [29]. The decision of deleting the cell data is taken by the forget gate and the input gate decides which information to be updated. The output gate gives the final output for the designed network [30].

CHAPTER 3

METHODOLOGY

The methods and descriptions of all the algorithms and experiments used in the remaining part of this thesis are discussed in this chapter. The first chapter describes a detailed description of ANN and LSTM and the techniques used to train them. Then there is a summary of the EEG signals, the problems faced during the training of these architectures, and finally how our algorithms used to classify the four-class EEG signals using classifiers like ANN and LSTM.

3.1 Experimental Studies

The datasets used in this study include EEG recordings of many subjects in four-class motor imagery.

The four-class MI applies to motor imagery gestures for the right hand, left hand, tongue, and foot. Data collection IIIa [1] of BCI competition III comprises of the EEG recording of three normal subjects. Nine subjects from BCI competition IV, dataset IIa [2],[3] are included in the second dataset list. The Brain-Computer Interfaces Laboratory (BCI-Lab) was used to collect both of the datasets used in this study. All participants who took part in the experimental work were asked to carry out four different motor imagery tasks based on a symbolic cue shown to the participants. In the next section, there is an explanation of both datasets, their time scheme diagram, and the steps for preprocessing used in this work.

3.1.1 Dataset IIIa from BCI Competition

1) Paradigm

The experiment begins with sitting on a relaxing chair with the subject. Subjects were asked to carry out Imagery movements for all the classes which include left-hand motor imagery, right-hand motor imagery, foot motor imagery, and tongue motor imagery in response to a cue. The signs were shown randomly. With 40 trials each, there were at least 6 runs. In total, each run consists of 8 seconds. The first two seconds were blank and silent, cue (beep sound) signals the beginning of a trial at time = 2s.

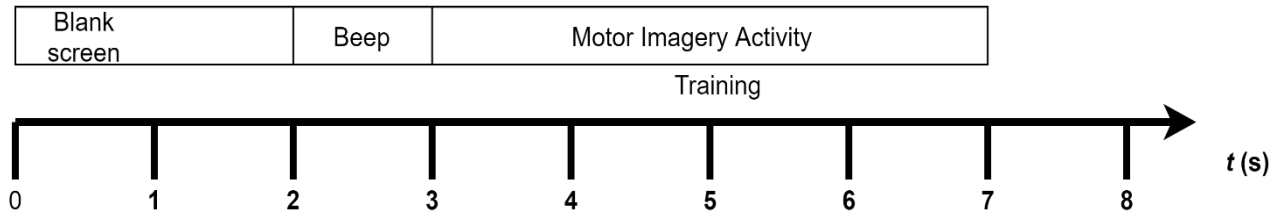


Figure 2: The time scheme for the paradigm

For 1s, an arrow symbol was displayed at time $t = 3s$ which points to the left, right, up, or down. In the meantime, the participant was told to perform imagery movement of the left hand, right hand, tongue, or foot according to the arrow displayed. This imagery process stops at $t = 7s$ before appearing the cross. The time scheme for this dataset is shown in figure 2. Participant 1 of this K3 experiment consisted of a total of 9 runs, resulting in a total of 360 trials for K3. For each class, there were 90 trials. Of the 360 trials, 180 were for training, and 180 were for evaluation purposes. There were 6 runs for the subject 2(K6) and subject 3 (L1), which resulted in 240 trials for each subject.

For the first two participants, there were 60 trials for each class. Data from the training and assessment comprises 120 trials for each subject with 30 trials for each class. As shown in figure 3, the raw signal is for a single subject. It contains all the data from a subject's trials. A channel or electrode placement on the scalp is defined by each color in the figure. It contains all the data from a subject's trials. In figure 3(a), the first 20 channels are shown in different colors and the remaining 21-40 and 41-60 channels are represented in figure 3(b) and figure 3(c) in the same way. The details of the labels are given in table 1.

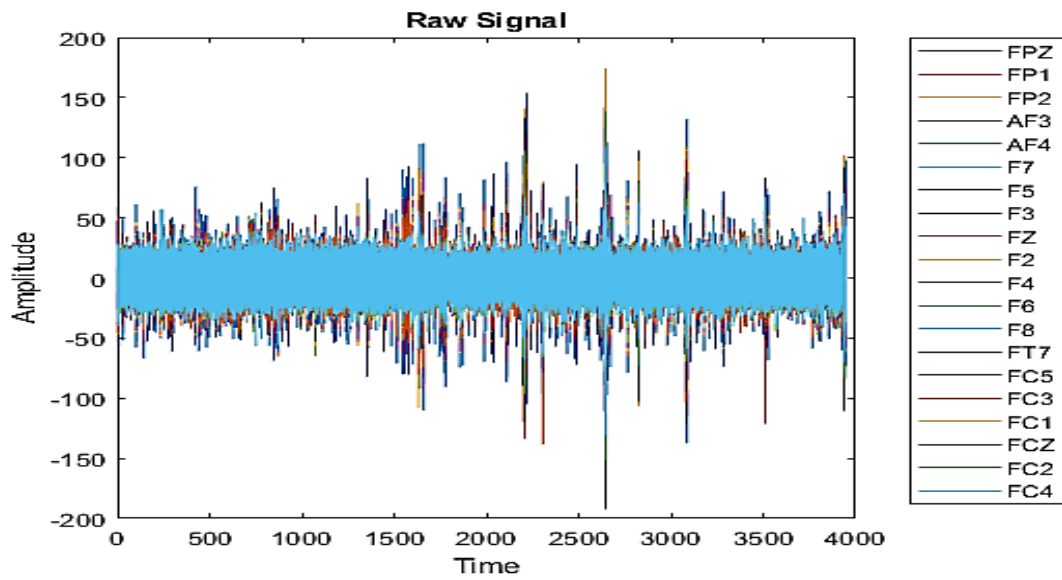


Figure 3(a): Raw signal part 1

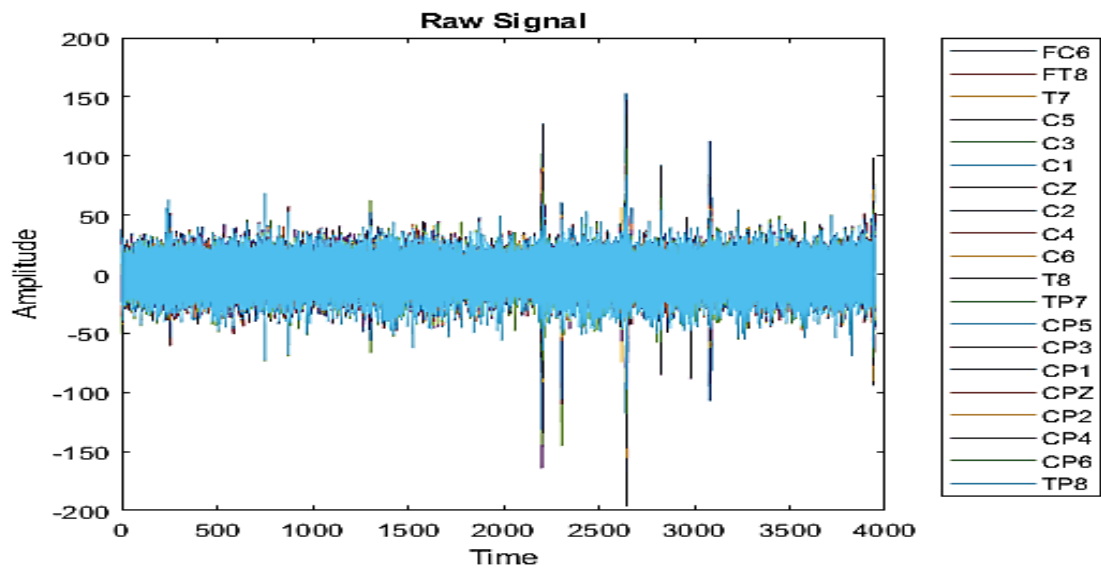


Figure 3(b): Raw signal part 2

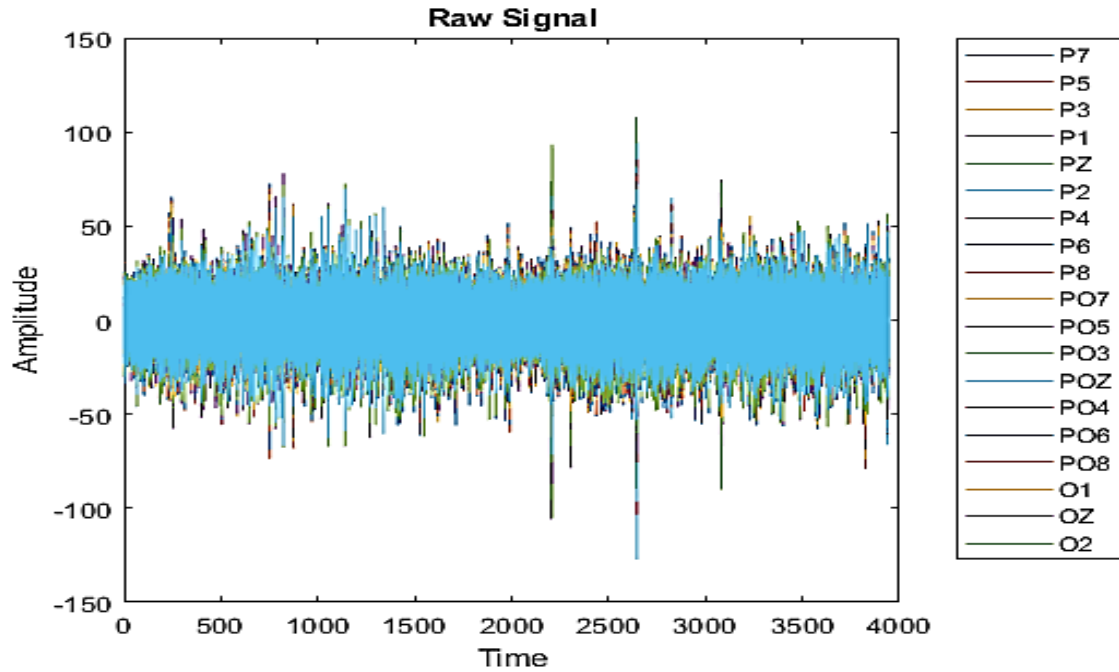


Figure 3(c): Raw signal part 3

Fp	Frontal polar
F	Frontal lobes
C	Central lobes
T	Temporal lobes
P	Parietal lobes
O	Occipital lobes
AF	Intermediate between Fp and F
FC	Intermediate between F and C
FT	Intermediate between F and T
CP	Intermediate between C and P
PO	Intermediate between P and O
TP	Intermediate between T and P
Z	Electrode position on the middle
Odd numbers	Electrode on the left hemisphere
Even number	Electrode on the right hemisphere

Table 1: Nomenclature

2) Pre-processing

A 64-channel EEG amplifier was used for the recording of the data of all three subjects; the left mastoid was used for comparison purposes while the right one mastoid was used as ground. Sixty EEG-based electrodes were mounted on the brain's scalp to record signals. The recorded EEG signals were sampled at 250 Hz. the data signals were filtered between 1 and 50 Hz to eliminate power line noise using a notch filter. Figure 3 displays the raw and filtered signals. The single-electrode raw signal is shown in figure 5(c) and the resulting signal is shown in figure 5(d) after applying the filter.

The artifacts containing trials were also included. The data trig value tells about the beginning of each trial and Class Label provided information about the classes marked "1" (left-hand motor imagery), "2" (right-hand motor imagery), "3" (tongue motor imagery), and "4" (foot motor imagery). Data from the first three seconds are omitted for our analysis and data from $t = 3s$ to $t = 7s$ is gathered where the motor imagery activity occurs.

Figures 4(a), 4(b), and 4(c), each with 20 channels, display a single trial for Class1. The trial contains the movement component of 4-sec motor imaging. It collects a sample of size 4s with 60 channels. At the sampling speed of 50 Hz, each sample is further segmented. A 64-channel EEG amplifier from neurocan was used to record data from all three subjects, the left mastoid was used as a reference and the right mastoid was used as a ground. Sixty EEG electrodes were mounted on the scalp of the brain to record signals. The recorded EEG signals were filtered between 1 and 50 Hz to eliminate power line noise.

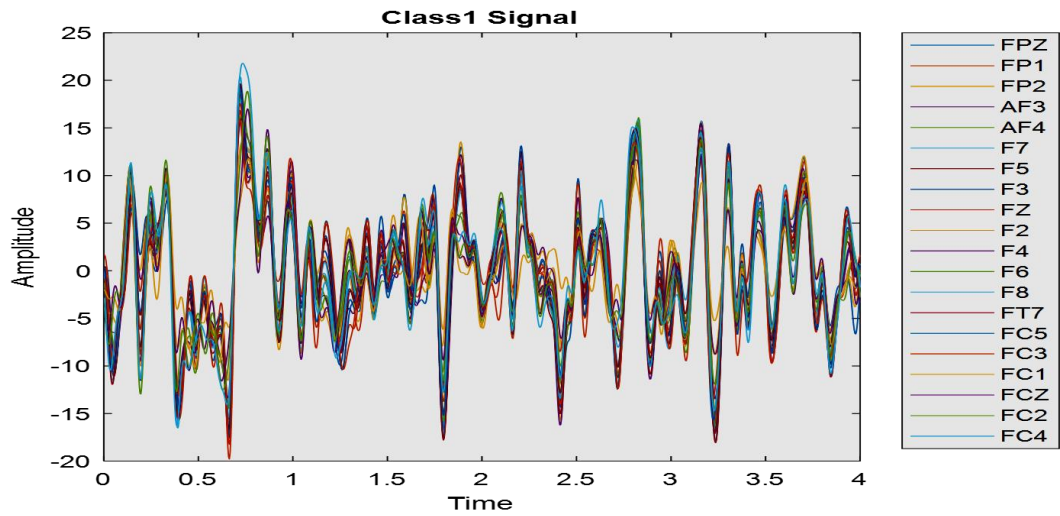


Figure 4(a): Single class trial part 1

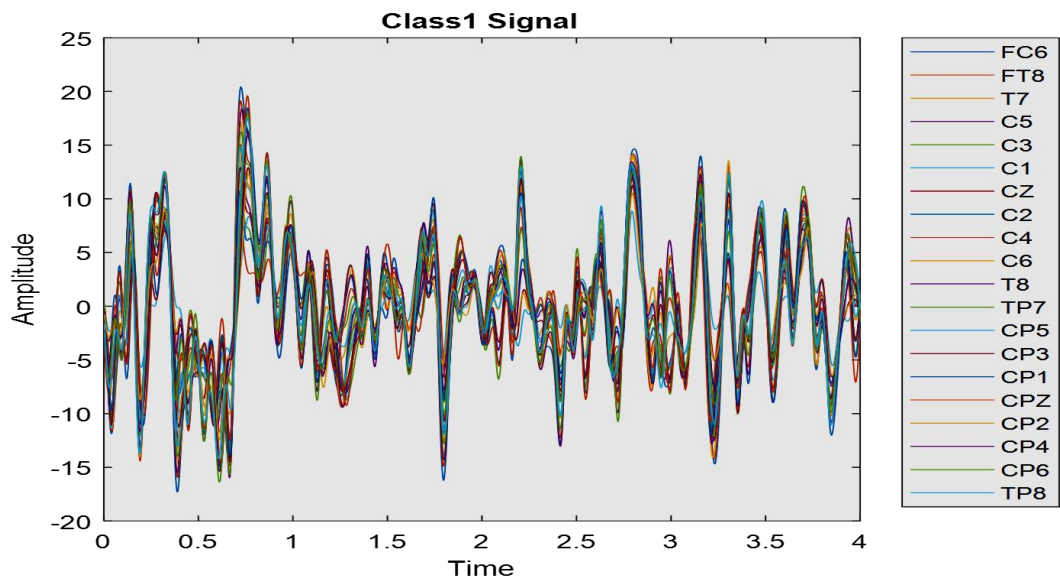


Figure 4(b): Single class trial part 2

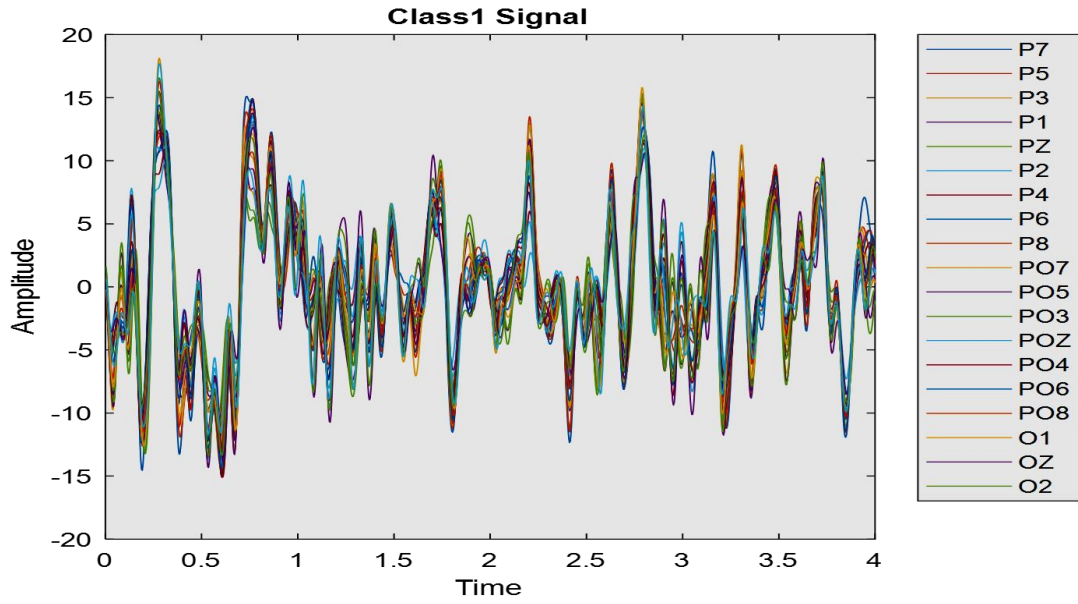


Figure 4(c): Single class trial part 3

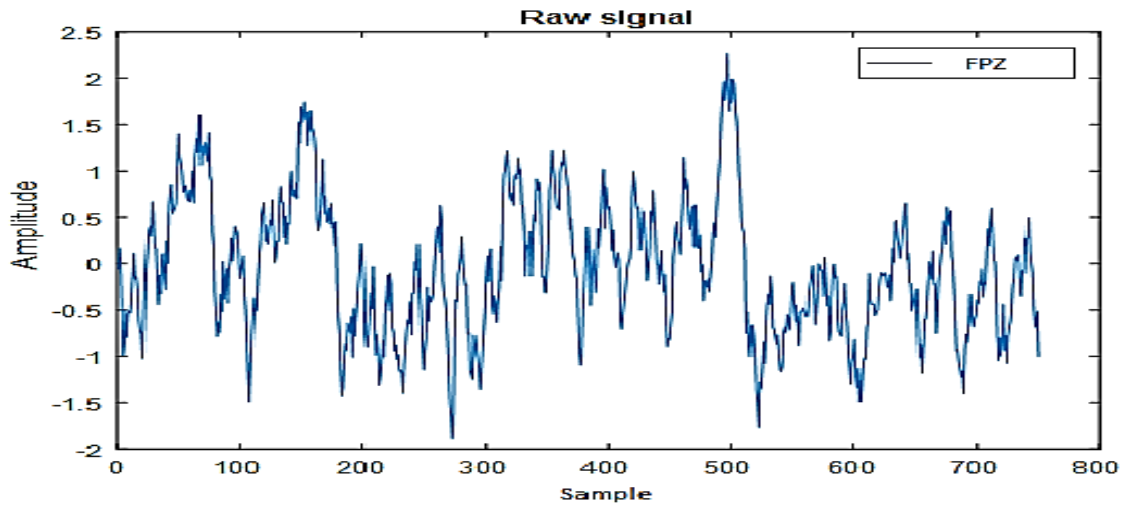


Figure 5(a): Raw signal

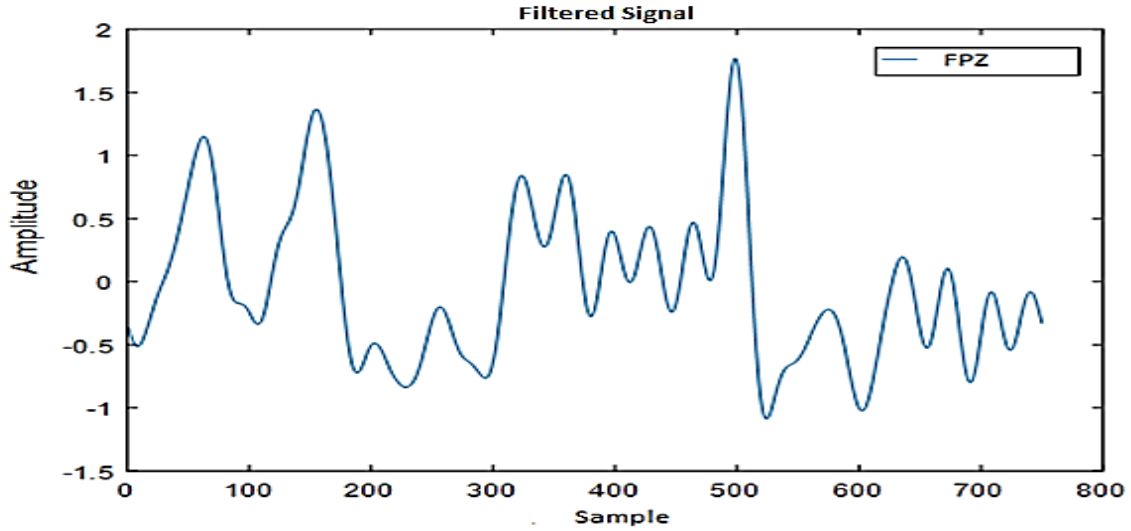


Figure 5(b): Filtered signal

3.1.2 Dataset IIa BCI Competition IV

1) Paradigm

The paradigm of data set IIa used in this work is shown in Fig. 3. At time $t = 0s$ an auditory signal along with a fixation cross appeared in a dark screen which lasted for 2s. A visual cue with an arrow pointing to the left, right, upward, and downward was presented at time $t = 2s$. These arrows act as a guide for all the subject's imagery movements of the left hand, right hand, tongue, or foot. The cue lasted for 1.25s. At time $t = 3s$ subjects started performing motor imagery task and it lasted for 3s. At time $t = 6s$ the motor imagery task was completed followed by a short break before starting the next trial. The complete scheme is shown in figure 6.

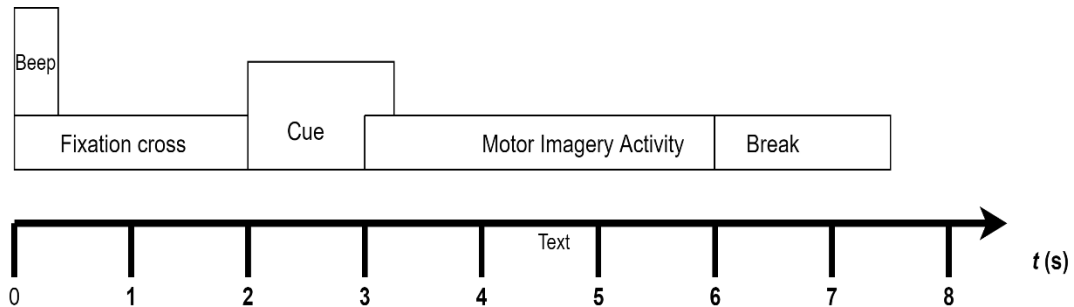


Figure 6: The time scheme for the paradigm

The data was collected from 9 subjects. There were four different classes labeled as '1', '2', '3', and '4'. Class 1 for left-hand motor imagery, 2 for a right hand, 3 for feet, and 4 for tongue imagery. The artifact data is included in this work. Two sessions were conducted on two different days for experimenting and there were a total of 6 runs in each of the sessions. Each run was separated by a short break. There were 48 trials in each run. Overall, there was 288 trial per session for each subject. For each class there were 72 trials, for training and evaluation, there were 288 trials for each subject. The raw data including all the runs along with short breaks for a single subject is shown in figure 7(a). Each color represents the channel or electrode placed on the scalp. There was a total of 25 channels. Since we were working on EEG signals only the first 22 channels were selected. Figure 7(b) shown single class motor imagery activity collected from time $t=3s$ to $t=6s$.

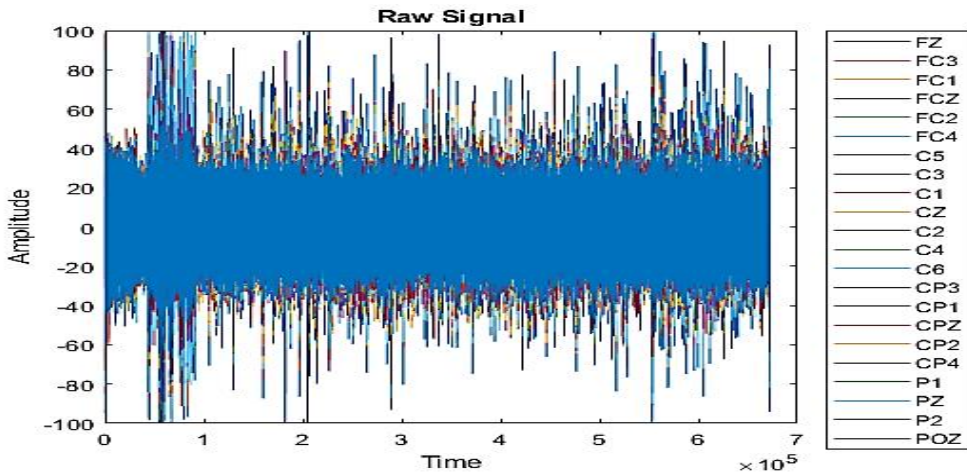


Figure 7(a): Raw signal of one subject's data

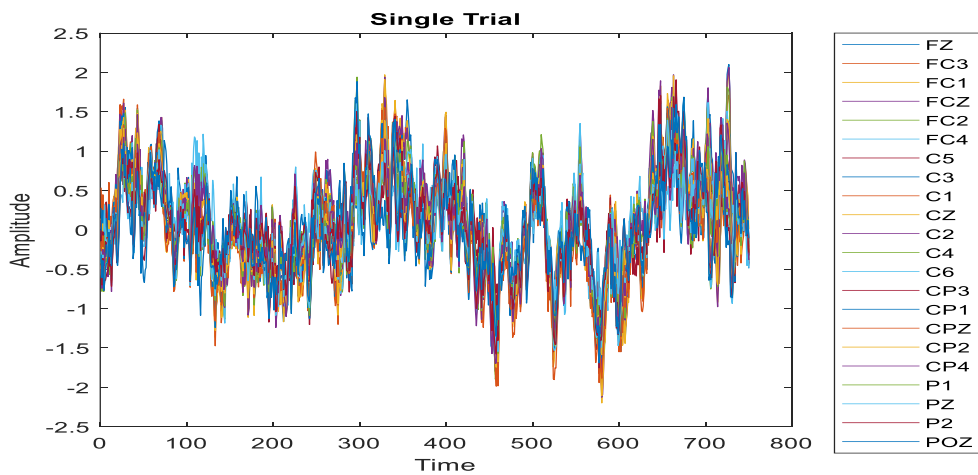


Figure 7(b): Single class trial

2) Pre-Processing

The experiment consisted of 9 participants and consisted of 25 channels in the data collection. There were 22 EEG channels and 3 monopolar electrooculograms (EOG) channels among those 25. As a reference, the left mastoid was used and the right mastoid was used as ground. 22 EEG and 3 EOG electrodes were mounted on the scalp of the brain to monitor signals, and the data signals collected were sampled at 250 Hz. using notch filters, the data signals were filtered between 1 and 50 Hz to remove power line noise again. Then a bandpass filter was implemented at a lower cut-off frequency of 0.5 Hz and a higher cut-off at 100 Hz. Only samples of data from 22 EEG channels were used in this work.

The raw data set was presented in GDF format and the BioSig toolbox functions were used for data loading [20]. We collected the data samples after the onset of cross fixation in a time interval from 3s to 6s. Each sample collected was further segmented at a 50 Hz rate. The data set is used to validate the classifier's accuracy results with a window size of 1s, 2s, and 3s. for training and testing, separate datasets were provided

3.2 Adopted Methodology

The approach that was implemented for this work is shown in figure 8. In the previous portion, data acquisition and data preprocessing were addressed. The next step is selecting the architecture after collecting the preprocessed signals. We have used ANN and LSTM architecture to classify the datasets for motor imagery used in this work. We trained the model after selecting the required architecture. The test datasets were then transferred to the trained model to obtain the accuracy results for both architectures. In later pages, the architectures used in this research and the precision outcomes are discussed.

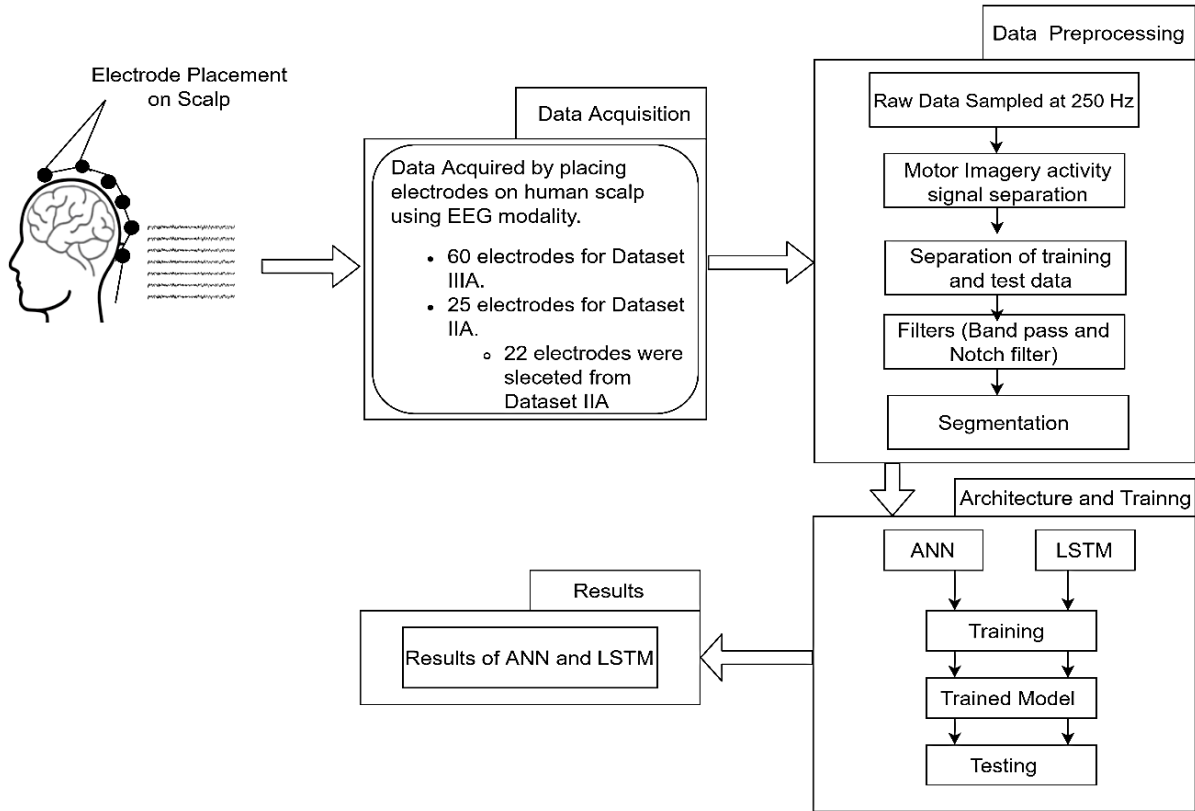


Figure 8: Adopted methodology flow chart

3.3 Deep Learning Method

Several layers of neurons are stacked above one another in deep neural networks. By adding hidden layers to the networks [21], the performance of the network can be increased. By increasing the number of layers, the complexity of the network rises. There is no need to manually extract features that are a challenging job in the machine learning approach, the key benefit of deep learning algorithms. We only have to feed the dataset to the network, which learns the characteristics automatically. In this work, we worked with LSTM in RNN and ANN architecture.

3.3.1 Artificial Neural Network (ANN)

ANN is a network of feed-forward neurons comprising various layers [22]. To update the weights, backpropagation is used. The architecture consists of 10 layers. In this architecture, there are various types of layers: an input layer, rectified linear units (ReLU) layer, dense layers which are also called fully connected layers. The input layer of the network is then transformed into a one-dimensional array and connected to eight hidden fully connected dense layers. In each hidden layer, the number of a neuron depends on input in every layer. Every input and output is linked via learnable weights in these layers. Non-linear activation function such as ReLU is also passed through each connected layer. Usually, the final flatten layer has the same number of nodes as the number of groups to which the input data must be classified. The activation function of the last layer is very carefully chosen and generally differ from previous fully connected layers. SoftMax, which normalizes outcomes between 0 and 1 based on the likelihood of each class, is one of the most commonly used layers. Usually, through stochastic gradient descent, neural networks are modified: Stochastic gradient descent was selected as the training algorithm (sgd). The batch size and layers were set at 64. By training the network repeatedly to obtain maximum accuracy, all hyperparameters such as the number of neurons, learning rate, weights, optimizer, and batch size were chosen empirically. The accuracy results of the selection of hyper-parameters used in this analysis are shown in table 2. It is easy to freely change the iterations of training [24]. In our work, 150 epochs of the model were prepared. The learning rate of 0.01 was chosen for this study. A higher learning rate causes the network to diverge inversely, and sluggish convergence is caused by a lower learning rate. The mean classification accuracy with different window sizes on the subjects was considered for the performance analysis metrics.

Table 2: Hyper-parameters for ANN

Number of layers	10
Number of Neurons	3
Learning rate	0.01
Activation function	Relu, Softmax
Loss function	Categorical cross-entropy
Maximum number of Epochs to train	150
Optimizer	Stochastic Gradient Descent (SGD)

3.3.2 Recurrent Neural Network (LSTM)

Long Short-Term Memory, one of the Recurrent Neural Network types, is the other architecture we use to explore the classification of EEG signals. Initially, the Recurrent Neural Network was represented by the study group of Schmidhuber and won eight international competitions in pattern recognition and machine learning [25],[26]. LSTM network provides the best results for data classification of time series and LSTM provides better results compared to traditional methods according to previous studies [27][28]. It operates on the concept of saving a layer's output and feeding that output to the next input layer, which assists in predicting the outcome of the layer. LSTM is one of the RNN types capable of learning from observation, which is an advantage of LSTM over other types of RNNs and neural networks. The key thing about LSTM is the cell state that can be modified, i.e. it can be removed or added. In LSTM, there are three gates: the forget gate, the input gate, and the output gate [29]. The decision of deleting the cell data is taken by the forget gate and the input gate decides which information to be updated. The final output of the network is given at the end of the output gate [30].

In this architecture, there are numerous layer types: the input layer, LSTM layer, rectified linear unit (ReLU) layer, dense layers are completely connected. For the LSTM layer pre-processed samples are given as input. The hidden layers that include the flatten layer and dense layers are connected to the LSTM layer. After each dense layer, the ReLU layer is added. The output layer that has SoftMax as the activation feature is followed by Relu. As four groups exist in our datasets, the output layer has 4 nodes. Adam was the optimizer used in LSTM. For the performance appraisal metrics, the mean accuracy of the classification of the subjects for

various window sizes was taken into account. The parameters used for LSTM architecture in this study are given in table 3. After training the network repeatedly, all the hyper-parameters were empirically selected to achieve maximum results for accuracy. The summary of LSTM is shown in figure 9.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 50, 60)	29280
lstm_2 (LSTM)	(None, 50, 60)	29040
flatten_3 (Flatten)	(None, 3000)	0
dense_4 (Dense)	(None, 32)	96032
dense_5 (Dense)	(None, 4)	132
Total params: 154,484		
Trainable params: 154,484		
Non-trainable params: 0		

Figure 9: Summary of LSTM

Table 3: Hyper-parameters for LSTM

Total no. of layers	5
Hidden layers	3
Learning rate	0.001
Activation function	Relu
Loss function	Categorical cross-entropy
Epochs	150
Optimizer	Adaptive moment estimation (ADAM)

RESULTS

To evaluate the performance of the architectures the trained models were tested against the test data provided in the dataset and the classification accuracy of Artificial Neural network (ANN) and Long-Short Term Memory (LSTM) of Recurrent Neural Network (RNN) was calculated against different window sizes. The test accuracy results using different window sizes were:

4.1 Long-Short Term Memory (LSTM) Architecture Results on Data Set IIIa :

The Model for Long-Short Term Memory on data set IIIa was trained using window sizes of 4s, 3s, 2s, and 1s respectively. The highest train and test accuracies were achieved while using the window of 4 sec and the average test accuracy was 96.99%. The result of all window sizes are shown below:

4.1.1 LSTM Accuracy and loss graph with 4sec Window:

The average accuracy achieved using window size 4sec was 0.9620. The graph of train and test accuracies is given in Figure 10(a). The accuracy is increasing gradually and after a certain number of epochs the highest accuracy is achieved and after that the graph line becomes straight. Here the green color shows the training accuracy and the blue line shows the test accuracy.



Figure 10(a): Accuracy graph

The loss graphs are for LSTM architecture using 4-sec window size is shown in Figure 10(b). The loss values are decreasing as the number of epochs are increasing and weights are updated and optimized. After a certain time, the loss values achieve their minimum values.

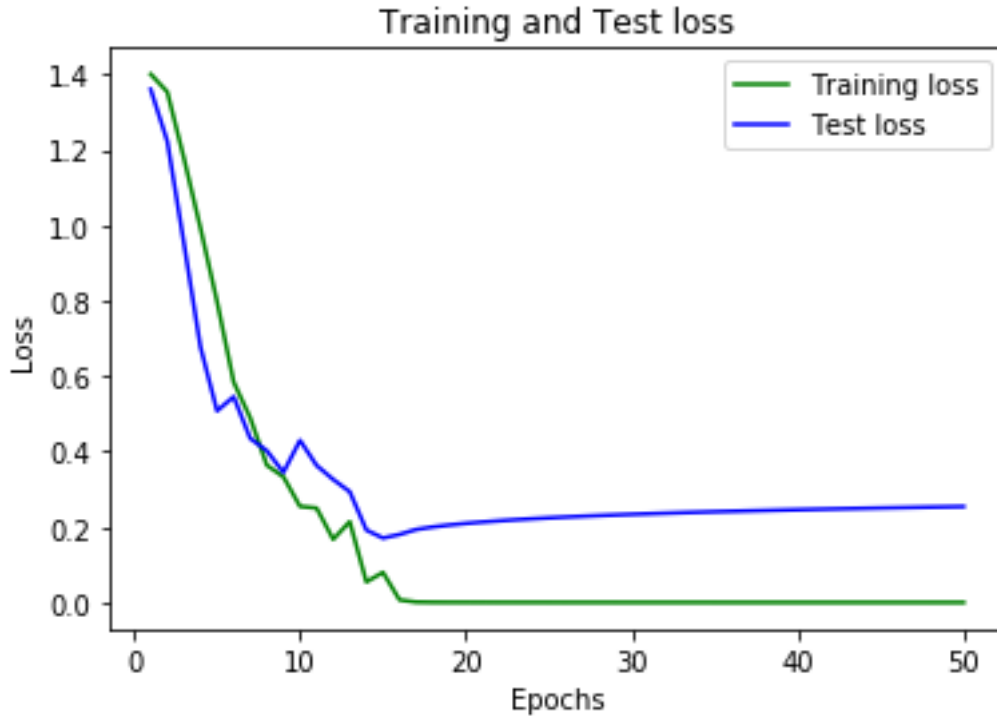


Figure 10(b): Loss graph

4.2 Artificial Neural Network (ANN) Results on Data Set IIIa :

The Model for Artificial Neural Network on data set IIIa was trained using window sizes of 4s, 3s, 2s, and 1s respectively. The highest train and test accuracies were achieved while using a window of 4 sec and the average test accuracy was 0.947%. The result of all window sizes are shown below:

4.2.1 ANN Accuracy and Loss graph with 4sec Window:

The average accuracy achieved using window size 4sec was 0.9470. The graph of train and test accuracies is given in Figure 11(a). The accuracy is increasing gradually and after a certain number of epochs the highest accuracy is achieved and after that the graph line becomes straight. Here the green color shows the training accuracy and the blue line shows the test accuracy.

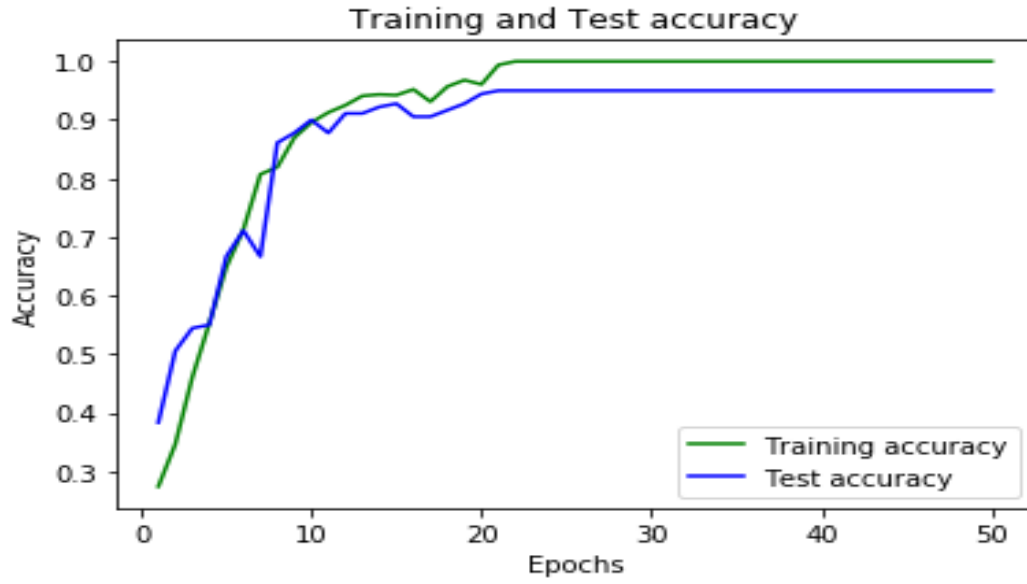


Figure 11(a): Accuracy graph

The loss graphs are for ANN architecture using 4-sec window size is shown in Figure 11(b). The loss values are decreasing as the number of epochs are increasing and weights are updated and optimized. After a certain time, the loss values achieve their minimum values.

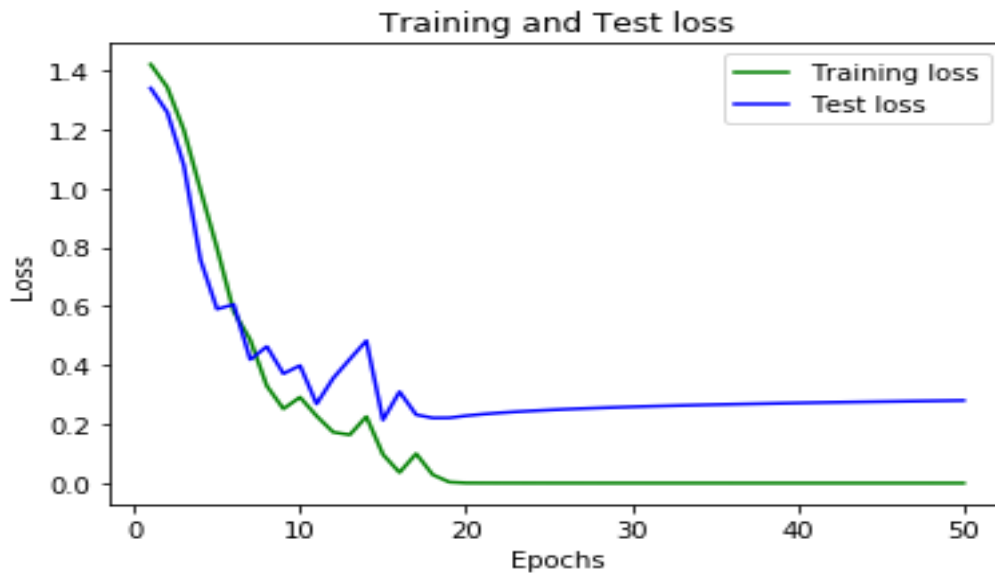


Figure 11(b): Loss graph

Table 4: Classification accuracies for dataset IIIa with different window sizes

Architecture	Window Size	K3	K6	L1	Average
ANN	4s	0.961	0.941	0.946	0.947
ANN	3s	0.933	0.921	0.929	0.927
ANN	2s	0.920	0.911	0.915	0.915
ANN	1s	0.85	0.82	0.840	0.830
LSTM	4s	0.960	0.961	0.952	0.9620
LSTM	3s	0.947	0.948	0.955	0.949
LSTM	2s	0.921	0.920	0.929	0.925
LSTM	1s	0.88	0.870	0.876	0.870

4.3 Long-Short Term Memory (LSTM) Architecture Results on Data Set IIa :

The Model for Long-Short Term Memory on data set IIa was trained using window sizes of 3s, 2s, and 1s respectively. The highest train and test accuracies were achieved while using a window of 3 sec and the average test accuracy was 0.945%. The result of all window sizes are shown below:

4.3.1 LSTM Accuracy and loss with 3sec Window:

The average accuracy achieved using window size 3sec was 0.9450. The graph of train and test accuracies is given in the figure. The accuracy is increasing gradually and after a certain number of epochs the highest accuracy is achieved and after that the graph line becomes straight. Here the green color shows the training accuracy and the blue line shows the test accuracy.

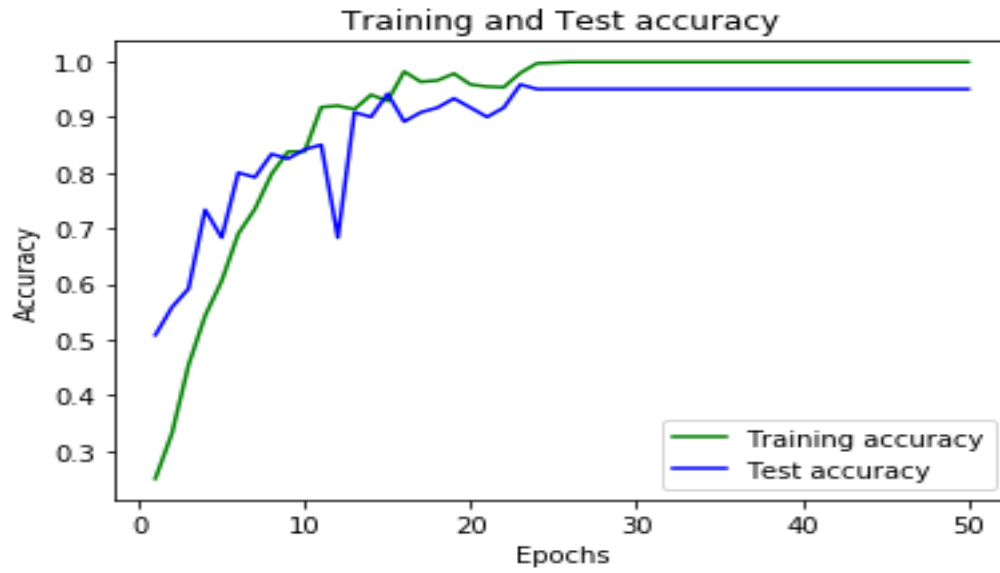


Figure 11(a): Accuracy graph



Figure 12(b): Loss graph

4.4 Artificial Neural Network (ANN) Results on Data Set IIa :

The Model for Artificial Neural Network on data set IIIa was trained using window sizes of 3s, 2s, and 1s respectively. The highest train and test accuracies were achieved while using a window of 3sec and the average test accuracy was 0.918%. The result of all window sizes are shown below:

4.4.1 ANN Accuracy and Loss with 3sec Window:

The average accuracy achieved using window size 3sec was 0.918. The graph of train and test accuracies is given below:

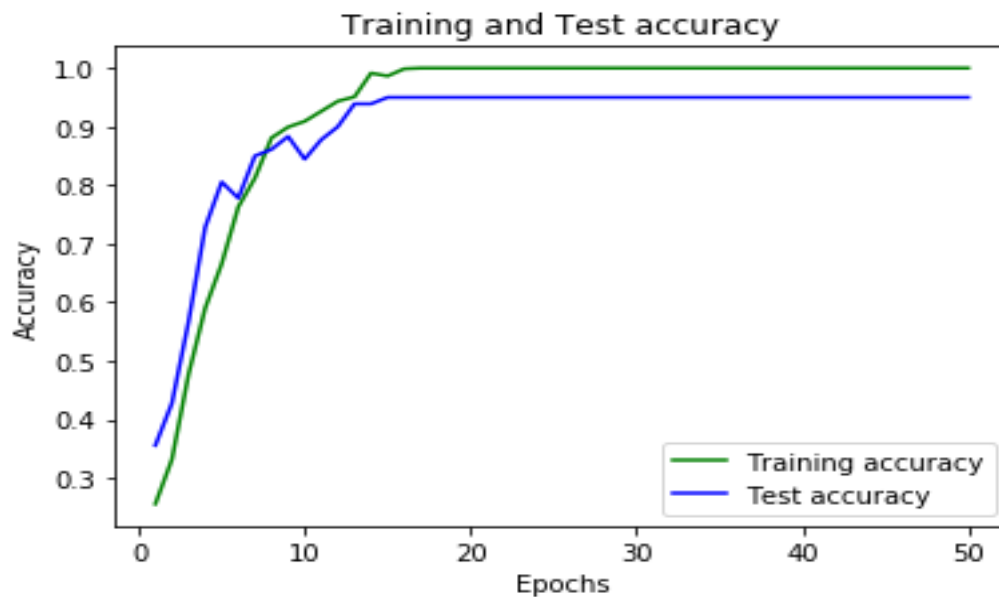


Figure 11(a): Accuracy graph

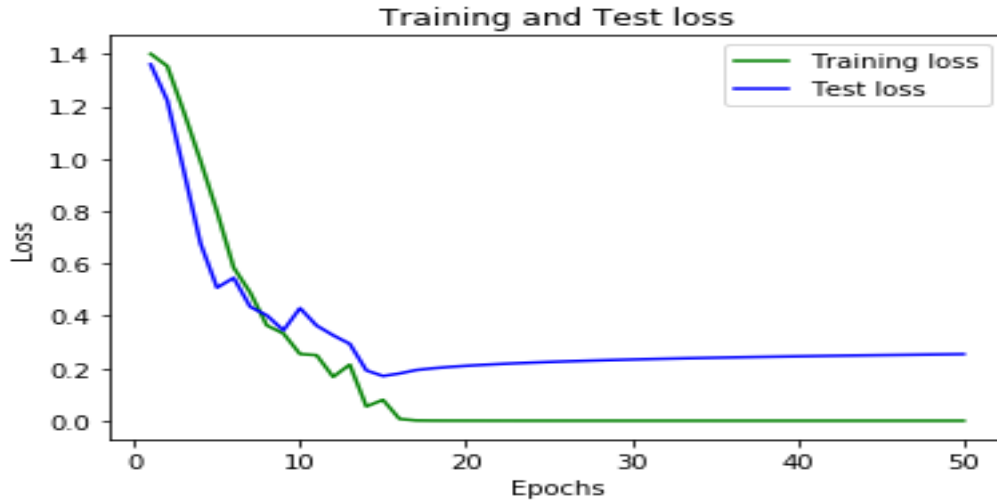


Figure 13(b): Loss graph

Table 4: Classification accuracies for dataset IIa with different window sizes

Architecture	Window Size	A1	A2	A3	A4	A5	A6	A7	A8	A9	Avg Accuracy
LSTM	3s	0.947	.933	.949	.945	.946	.947	.949	.948	.947	.945
LSTM	2s	.921	.915	.923	.918	.920	.919	.923	.921	.911	.919
LSTM	1s	.845	.838	.847	.840	.843	.841	.844	.845	.838	.842
ANN	3s	.919	.916	.920	.919	.920	.9175	.922	.919	.918	.918
ANN	2s	.867	.859	.871	.865	.866	.878	.873	.876	.865	.868
ANN	1s	.828	.817	.825	.823	.824	.826	.827	.826	.819	.823

CONCLUSION

5.1 CONCLUSION AND DISCUSSION:

In this study, we examined the classification algorithms for four EEG signals based on class motor imagery. The objective of the study was to analyze and evaluate the efficiency of the machine and the technique of deep learning models critically. To train, validate, and evaluate the deep learning models, two publicly accessible motor imagery data sets comprising of a total of 13 subjects and 4 classes were used. The data sets were in the form of electrical signals, and the data was pre-processed before training the model. To classify the datasets, ANN algorithms were used. To test the performance of both the classifiers, classification accuracies have been reported. The impact of choosing the various size of windows is also measured. The findings indicate that increasing the size of the window has a better effect on the accuracy of classification.

For dataset IIIa, BCI competition IIIa comparison in the table is shown according to a study. For accuracy results of this dataset, our proposed LSTM gives the highest performance having an average accuracy of 0.962. The second one, with a precision of 0.947, is given by ANN. The least precision algorithm with the least accuracy is LDA. BCI competition IV results for dataset IIa are identical to the previous dataset, with LSTM achieving the highest score with an accuracy rate of .9455. With an accuracy of 0.918, ANN is the second-high accuracy giving classifier, while KNN has the lowest average accuracy of 0.679.

In addition to manual extraction of features, machine learning algorithms prefer to generalize complex data patterns, resulting in poor results when the number of classes is increased. Most previous studies have used standard machine learning algorithms to operate on binary groups. The accuracy achieved for the classification of two mental tasks was 87 percent using the traditional ML algorithm SVM [32]. In the same analysis, using SVM on two separate brain signals, 87.2 percent average accuracy was calculated.

Table 6. Comparison of the different machine learning algorithms with deep learning algorithms on dataset IIIa, BCI competition III, and dataset IIa, BCI competition IV.

Approach	Year	Dataset	Average Accuracy
LDA	2018	BCI IIIa	0.803
KNN	2018	BCI IIIa	0.786
SVM	2018	BCI IIIa	0.793
NB	2018	BCI IIIa	0.810
FLS	2018	BCI IIIa	0.865
ANN (Proposed)	2020	BCI IIIa	0.947
LSTM (Proposed)	2020	BCI IIIa	0.962
LDA	2018	BCI IIa	0.712
KNN	2018	BCI IIa	0.679
SVM	2018	BCI IIa	0.712
NB	2018	BCI IIa	0.704
FLS	2018	BCI IIa	0.726
ANN (Proposed)	2020	BCI IIa	0.918
LSTM (Proposed)	2020	BCI IIa	0.945

The findings have shown that in classifying the datasets, LSTM outperforms ANN. For dataset III, on four different window sizes, the precision results of ANN vary from 83.0 to 94.7 percent, and for LSTM it ranges from 87.1 to 96.2 percent for a window size from 1s to 4s. the accuracy for ANN ranges from 82.3 to 91.8 percent for dataset IIa, and it ranges from 84.2 to 94.5 percent for LSTM for a window size of 1s to 3s.

REFERENCES:

- [1] B. Blankertz et al, “ The BCI competition III: Validating alternative approaches to actual problems,” *IEEE Trans. Neural Syst. Rehabil.Eng.*, vol. 14, no. 2,pp. 153_159,jun. 2006.
- [2] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, “Filter bank common spatial pattern algorithm on BCI Competition IV Dataset 2a and 2b ”, *Frontier Neurosci*, vol. 6, p.39, Mar. 2012.
- [3] R.A. Ramadan and A. V. Vasilakos, “Brain-computer interface: control signals review”, *Neurocomputing*, vol. 223, pp. 26-44, 2017.
- [4] J. van Erp, F. Lotte, and M. Tangermann, “Brain-computer interfaces: Beyond medical applications,” *Computer*, vol. 45, no. 4, pp. 26_34, 2012.
- [5] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain-computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211_1279, 2012.
- [6] M. Ahn, M. Lee, J. Choi, and S. C. Jun, “A review of brain-computer interface games and an opinion survey from researchers, developers, and users,” *Sensors*, vol. 14, no. 8, pp. 14601_14633, 2014.
- [7] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, “Brain-computer interfacing: Applications and challenges,” *Egyptian Informatics Journal*, vol. 16, no. 2, pp. 213–230, 2015.
- [8] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, AB, 2017, pp. 234-239, 2017.
- [9] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [10] H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil, “Multilayer perceptron: Architecture optimization and training.,” *IJIMAI*, vol. 4, no. 1, pp. 26–30, 2016.
- [11] D. F. Morrison, *Multivariate analysis, overview*. Wiley Online Library, 1998.
- [12] J. Yang, H. Singh, E. L. Hines, et al., “Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach,” *Artificial Intelligence in Medicine*, vol. 55, no. 2, pp. 117–126, 2012.
- [13] J. M. Aguilar, J. Castillo, and D. Elias, “EEG signals processing based on fractal dimension features and classified by neural network and support vector machine in motor imagery for a BCI,” in *VI Latin American Congress on Biomedical Engineering CLAIB 2014*, Paraná, Argentina 29, 30 & 31 October 2014, A. Braidot and A. Hadad, Eds., vol. 49 of *IFMBE Proceedings*, pp. 615–618, Springer International Publishing, 2015.
- [14] M. Serdar Bascil, A. Y. Tesneli, and F. Temurtas, “Multichannel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for the brain-computer interface,” *Australasian Physical & Engineering Sciences in Medicine*, vol. 38, no. 2, pp. 229–239, 2015.
- [15] R. J. Williams and D. Zipser, “A learning algorithm for continually running fully recurrent neural networks,” *Neural computation*, vol. 1, no. 2, pp. 270–280, 1989.
- [16] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [17] B. Blankertz et al., "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.
- [18] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.
- [19] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.
- [20] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain-computer interfaces, a review," *Sensors*. 2012.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [22] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10-year update," *J. Neural Eng.* 15,031005, 2018.
- [23] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, pp. 464–472., 2017.
- [24] Y. Bengio "Practical Recommendations for Gradient-Based Training of Deep Architectures", In G.Montavon, G.B. Orr, KR. Müller (eds) *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, vol. 7700, 2012.
- [25] J. Schmidhuber and S. Hochreiter. "Long short-term memory". *Neural Computation*, vol. 9, pp. 1735–1780, 1997.
- [26] L. C. Schudlo, and T. Chau, "Dynamic topographical pattern classification of multichannel prefrontal NIRS signals":
 II. Online differentiation of mental arithmetic and rest. *J. Neural Eng.*, vol. 11:016003, 2013.
- [27] A. Graves, M. Liwick, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A Novel connectionist system for unconstrained handwriting recognition". *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 31, pp. 855–868, 2009.
- [28] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Vancouver, BC, 2013.
- [29] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey". *IEEE Trans. Neural Network. Learn. Syst.*, vol. 28, pp. 2222–2232, 2016.
- [30] Y. Luan and S. Lin, "Research on Text Classification Based on CNN and LSTM," *2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, Dalian, China, 2019, pp. 352–355, 2019.
- [31] T. Nguyen, I. Hettiarachchi, A. Khatami, L. Gordon-Brown, C. P. Lim and S. Nahavandi, "Classification of Multi-Class BCI Data by Common Spatial Pattern and Fuzzy System," *IEEE Access*, vol. 6, pp. 27873–27884, 2018.
- [32] N. Naseer, and K.-S. Hong., "Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a brain-computer interface", *Neurosci. Lett.*, vol. 553, pp. 84–89, 2013.