ALCOHOLIC AND NON-ALCOHOLIC EEG SIGNALS CLASSIFICATION USING DEEP LEARNING METHODS

A PROJECT REPORT

submitted by

JITHIN SAMUEL
(Reg. No. TKM20EEII12)

to

the APJ Abdul Kalam Technological University
in partial fulfillment of the requirements for the award of the Degree

of

Master of Technology
in
Electrical and Electronics Engineering

with specialisation in

Industrial Instrumentation and Control

Department of Electrical and Electronics Engineering
TKM College of Engineering
Kollam - 691005
KERALA

JULY 2022

DECLARATION

I undersigned hereby declare that the project report entitled "**Alcoholic and Non-alcoholic EEG Signals Classification Using Deep Learning Methods** ", submitted for partial fulfillment of the requirements for the award of degree of Master of Technology in Electrical and Electronics Engineering with specialisation in Control Systems, of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of *Dr.Sabeena Beevi K* , Associate Professor, Head of the Department, Department of Electrical and Electronics Engineering,TKM College of Engineering. This submission represents my ideas in my own words and where ideas or words of others have been included. I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Kollam
July 01 , 2022

JITHIN SAMUEL

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM - 691005

CERTIFICATE

This is to certify that the report entitled " **Alcoholic and Non-Alcoholic EEG Signals Classification Using Deep Learning Methods** " submitted by **JITHIN SAMUEL** , **(Reg. No. TKM20EEII12)** of fourth semester to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Electrical and Electronics Engineering with specialisation in Industrial Instrumentation and Control, is a bonafide record of the project work done by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Dr.Sabeena Beevi K
Project Supervisor
Associate Professor
Department of EEE
TKM College of Engineering

Prof.Sumayya Jaleel
Project Coordinator
Assistant Professor
Department of EEE
TKM College of Engineering

Prof.Shanavas TN
PG Coordinator
Associate Professor
Department of EEE
TKM College of Engineering

Dr.Sabeena Beevi K
Head of the Department
Associate Professor
Department of EEE
TKM College of Engineering

Acknowledgment

A lot of effort and hard work has been put into this project in course of its presentation. However, it would not have been possible without the kind support and help of many individuals and other sources. I would like to extend my sincere thanks to all of them. I take this opportunity to express my deep sense of gratitude and sincere thanks to all who helped me to complete this project report successfully.

I express my sincere thanks to *Dr.T A Shahul Hameed*,Principal,TKM College of Engineering for his encouragement in the completion of my project.

I thank *Dr.Sabeena Beevi K*,Associate Professor, Head of the Department, Department of Electrical and Electonics Engineeering, *Dr.Imthias Ahamed TP*,Head of Self-financing programs and *Prof.Shanavas T N*, PG Coordinator,Department of Electrical and Electonics Engineeering, for their support and cooperation.

I am greatly thankful to my Guide *Dr.Sabeena Beevi K* , Associate Professor, Head of the Department, Department of Electrical and Electonics Engineeering for her supervision,assistance and helpful suggestions.

Iam deeply indebted to Project Coordinator *Prof.Sumayya Jaleel*, Assistant Professor,Department of Electrical and Electonics Engineeering and *Prof.Amal A*, Associate Professor,Department of Electrical and Electonics Engineeering,for their excellent guidance,positive criticism and valuable comments.

Finally I thank my parents and friends near and dear ones who directly and indirectly contributed to the successful completion of my project project.

JITHIN SAMUEL

Abstract

Alcohol has been ranked as one of the five most addictive substances. Alcoholism is a critical disorder that affects the one using it. Electroencephalography(EEG) is a puissant method for detecting alcoholism. Earlier, majority of studies for alcoholism detection were comprised of various deep learning and machine learning techniques, but they cannot takeout deep lurked characteristics of EEG. Hence, present a deep learning model that helps to automatically classify both alcoholic and non-alcoholic EEG signals most accurately. To scrutinize this, this work proposed one machine learning based algorithms, three deep learning based algorithms and one ensembled model for grouping of alcoholic and non-alcoholic EEG data for comparison. Existing deep learning models was based on ANN and LSTM with highest classification accuracy of 92.74%. The machine learning approach applied in this study for classification is Random Forest. Principal Component Analysis(PCA) based feature minimization mechanism has been used to reduce dimension of the EEG data and then output of PCA are used as input to Random Forest model. Raw EEG data are straightly applied as input to ensembled models and other deep learning models that we use in this study for classification which are Gated Recurrent Unit(GRU), Bidirectional Long Short Term Memory(BiLSTM),Vanilla LSTM. All of the proposed techniques were evaluated using a publicly accessible UCI alcoholic EEG dataset. With the BiLSTM model we achieved the greatest accuracy of 95.87%.

Contents

List of Tables

List of Figures

Abbreviations

ML	Machine Learning
DL	Deep Learning
ANN	Artificial Neural Network
LSTM	Long Short Term Memory
EEG	Electroencephalogram
PCA	Principal Component Analysis
GRU	Gated Recurrent Unit
BiLSTM	Bidirectional LSTM
AI	Artificial Intelligence

Notations

b	bias contant
W	Weight of neurons
f_t	forget gate
x_t	Present input of LSTM
h_{t-1}	previous hidden state
t	time step
i_t	input gate of LSTM
v_t	value generated by tanh function
ct	cell state
c_{t-1}	previous cell state

Chapter 1

INTRODUCTION

1.1 Overview

Alcohol disrupts the brain's communication connections and can alter the way it appears and functions. Alcohol makes it more difficult for the parts of the brain that govern balance, memory, speech, and judgement to execute their tasks, increasing the risk of accidents and other bad consequences[1] . Long-term excessive drinking produces changes in the neurons, such as size reductions , disrupts the function of neurons in the central nervous system and changes a person's behaviour. Cardiomyopathy, stroke, high blood pressure, cirrhosis, and an increased risk of cancer are the most adverse health repercussions of heavy alcohol consumption.

Alcohol has a variety of effects on the body, but it is most harmful to the lungs, cortex, cardiac, and immune system . The World Health Organization (WHO) estimates that over 3.4 million people die each year as a result of binge drinking[2]. Short-term consumption of excessive alcohol use has immediate consequences, raising the risk of a variety of health problems,accidents, falls, drownings,burns,homicide, suicide, sexual assault,intimate relationship violence. Long-term consumption of excessive alcohol use can result in the development of chronic illnesses and other major health issues . Alcohol abusers are known to struggle with decision-making, loss of memory power, and behavioural changes[3]. Over time, excessive alcohol consumption may result in pancreas inflammation and pancreatitis. Abdominal pain and digestive enzyme release can both be brought on by pancreatitis. Chronic liver inflammation and alcohol-related liver damage are both made more risk by long-term alcohol usage. Alcohol-related liver disease is a potentially life threatening condition that leads to toxins and waste buildup in your body. Chronic liver inflammation can cause scarring, or cirrhosis. When scar tissue forms, it may permanently damage your liver. Drinking over time can also harm your frontal lobe, which is the area of the brain in charge of executive processes like abstract thought, judgment, social behavior, and performance. Wernicke-Korsakoff syndrome, a brain illness that affects memory, is among the lasting brain damage that can result from chronic excessive drinking[4].

The detection of alcohol in people is a challenging process since most devices rely on the scent of the alcohol, which is not always accurate[5].Smoking,mouth wash,breath fresheners can affect the results of this breath alcohol test. Electroencephalogram is a test that uses tiny metal discs (electrodes) connected to the scalp to assess electrical activity in the brain[21]. Brain cells use electric pulses to communicate and are constantly active, even while you are sleeping. On an EEG recording, this activity appears as wavy lines. EEG signals have been defined as a useful analysing tool in the identification of alcoholics. EEG signals from the scalp give a record of brain electrical activity. An EEG is used to either affirm or reject conditions like alcoholism in high sensational cases. In accident cases it's not easy to find out whether the victim is drunken or not,but in such cases EEG is a powerful for the identification of alcoholism. EEG recordings include enormous amounts of information with dynamic properties. The EEG data has thus far been visually analysed in order to identify and explain anomalies in the brain and how they spread. In adults, alcohol use seems to be linked to a certain arrangement of brain impulse activity[6]. This manual method of analysing large amounts of data is wasteful and inaccurate, it takes time and resources, and human error adds to lower decision-making accuracy. There is currently no viable method for automatically, quickly, and reliably detecting alcoholism using EEG data. As a result, there is a substantial need for an EEG data analyzer that is both automated and economical and can create credible scientific evidence that is accurate and up-to-date.

1.2 Objectives

The main objective of this work is to create deep learning model to reliably divide EEG signals into two groups, alcoholic and non-alcoholic as accurately as possible. The main contributions of this study are: (1) Design and validation of a new framework for the most accurate automatic identintification of alcoholic and non-alcoholic subjects. (2) Exploration of the significance of the application of the feature reduction method on applying a deep learning and machine learning algorithm. (3) Work towards a accurate classification scheme for the suggested features to separate the subjects.(4) Improvement of grouping accuracy compared to existing deep learning methods.

1.3 Organisation of the report

The previous works related to thesis are discussed in chapter 2.Chapter 3 deals with Science of machine learning and deep learning techniques. Previous implemented models are addressed in chapter 4.Chapter 5 describes the proposed models for alcoholic EEG classification.Experimental results are discussed in chapter 6 and the conclusion of the thesis is given in chapter 7.

Chapter 2

LITERATURE REVIEW

2.1 Overview

This chapter studies the previous works related to thesis such as the different deep learning and machine learning based alcoholic EEG signals classifiers.

2.2 Past works

Many studies based on diagnosing alcoholism using EEG signals have been carried out in recent years. Bajaj *et al.* [7] proposed a technique for figuring out alcohol-related EEG signals based on non-negative least squares classifier(NNLS) and the Short Time Fourier Transform (STFT). The purpose of this work was to present a reliable approach for automatically distinguishing alcoholic EEG signals based on time-frequency (T-F) picture data, which conveys important EEG signal properties. In this article, we suggest a novel hybrid technique for automatically classifying the EEG data from control and alcoholic subjects. The suggested method is based on texture picture feature extraction, time-frequency images, and a nonnegative least squares classifier (NNLS). The spectrogram of the short-time Fourier transform is taken into account in T-F analysis. After being collected, the T-F pictures are converted into 8-bit grayscale versions. The co-occurrence of Eig(Hess)-CoHOG features and the histograms of oriented gradients (CoHOG) is retrieved from T-F pictures. Finally, the collected features are used as input by the NNLS classifier to categorise alcoholic signals and regulate EEG signals.

Faust[8] described a technique in which multiple machine learning methods were combined with wavelet packet decomposition to derive energy measurements that could be used to identify alcoholic EEG data. Algorithms for feature extraction and classification were used to build the identification system. Wavelet packet decomposition (WPD) and energy metrics served as the foundation for the feature extraction. The statistical t-test approach was used to establish feature fitness. A competitive 10-fold cross-validated investigation of six classification algorithms used the retrieved characteristics as training and test data. This study found that the k-nearest neighbour (k-NN) approach outperformed the fuzzy Sugeno classifier (FSC), probabilistic neural network (PNN), Gaussian mixture model (GMM), and decision tree .

In [9] Acharya *et al.* described a technique for grouping alcoholic and normal EEG data using nonlinear characteristics giving to an SVM classifier. In [10] for categorization of alcoholic and control EEG data, auto regressive (AR) and the fast Fourier transform (FFT) approach based power density were utilised as characteristics with a machine learning strategy. Here, the PSD derived from normal, epileptic, and drunken EEG data was compared. Welch's approach,auto Walker's regressive (AR) method, and Burg's method were used to determine the power density spectrum.

For automatic alcoholic signal identification, Anuragi *et al.*[11] employed machine learning models focused on flexible analytical wavelets transform (FAWT). Here, automated alcoholism detection machine learning algorithms based on the flexible analytical wavelets transform (FAWT) are developed. The suggested technique uses FAWT to decompose EEG signals into approximative and precise wavelet coefficients. The chosen wavelet coefficients are used to extract statistical variables including mean, standard deviation, kurtosis, skewness, and Shannon entropy. The various machine learning models receive the features. In [12], employed correlation dimension-based characteristics as measurements to distinguish between intoxicated and non-alcoholic EEG data.

For the classification of normal and intoxicated EEG data , Kousarrizi *et al.*[13] used Wavelet Transform (WT) based features with Support vector method (SVM).Supriya *et al.*[14] developed a data evaluation strategy for distinguishing intoxicated EEG signals from control signals using a combination of complex network (CN) and some machine learning techniques (e.g., Naive Bayes (NB), Linear Discriminant Analysis (LDA). In [3] classification of alcoholic and non-alcoholic EEG Signals is based on Sliding Singular Spectrum Analysis (S-SSA) and Independent Component Analysis(ICA) ,here to identify alcoholic and non-alcoholic EEG data, these components were utilised to train and evaluate several machine learning models such as Support Vector Machine(SVM),K-Nearest Neighbor(KNN),Artificial Neural Network (ANN), Adaptive Boosting(AdaBoost) and Extreme Gradient Boosting (XGBoost). Alcoholic EEG signals classification based on deep learning algorithm Long Short-Term Memory(LSTM) is proposed in [16].

The vast bulk of research so far on alcoholism identification using EEGs has been done in the field of machine learning, as can be seen from the above literature review. Only study based on the LSTM algorithm was proposed in the deep learning sector for the categorization of intoxicated and control EEG data with an accuracy of 93%. The available techniques rely on feature extraction techniques (e.g. Fourier Transform, principal component analysis,Entropy) and classic machine learning algorithms (e.g. SVM, k-nearest neighbour) that are manually picked based on the researcher's deep idea. Existing feature extraction algorithms are unable to recover the EEG signals deep hidden characteristics from several layers. Previous machine learning algorithms are not able to identify aberrant elements from the deep hidden layer efficiently due to their single layer structure. Moreover existing deep learning method based on LSTM has an accuracy of only 92.74%.

2.3 Conclusion

This chapter discussed about the previous works related to the thesis. Most of existing method has been performed based on machine learning methods and deep learning models. They can't extract deep lurked characteristics of EEG signals. Existing methods have limitation in their efficiency and accuracy. The next chapter deals with the detailed studies about the Science behind machine learning and deep learning techniques.

Chapter 3

ARTIFICIAL INTELLIGENCE

3.1 Overview

John McCarthy first outlined the concept of artificial intelligence in 1956. He described AI as:‘The science and engineering of making intelligent machines’.

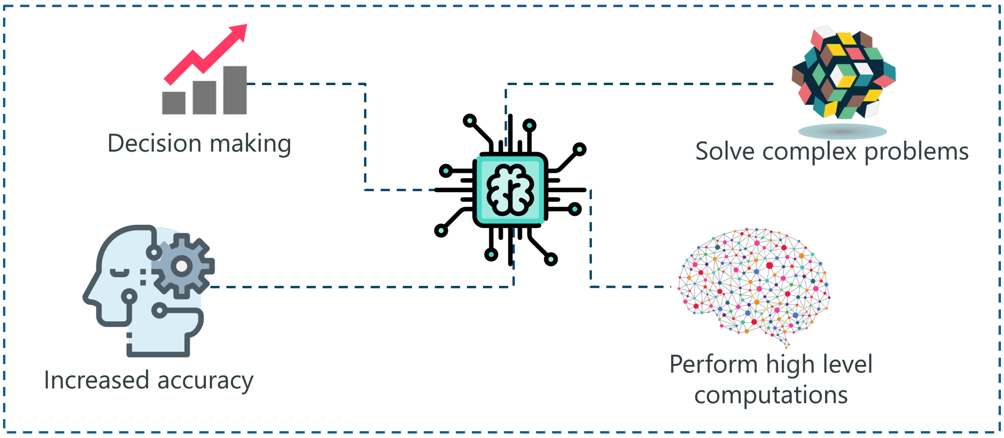

Figure 3.1: The science of Artificial Intelligence.

The creation of computer systems that can carry out tasks that need human intelligence, such as decision-making, object detection, solving complex issues, and so forth, is known as artificial intelligence. Vendors have been rushing to highlight how their goods and services integrate AI as the hype surrounding the AI technology has grown. Frequently, what they mean by AI is just one element of AI, like machine learning. For the creation and training of machine learning algorithms, AI needs a foundation of specialised hardware and software. There is no one programming language that is exclusively associated with AI, but a few including Python, R, and Java.

3.2 Branches Of Artificial Intelligence

By employing the following methods, artificial intelligence can be applied to resolve practical issues:

1. Machine Learning
2. Deep Learning
3. Natural Language Processing
4. Robotics
5. Expert Systems
6. Fuzzy Logic

3.2.1 Machine Learning

Machine Learning(ML) is one of the most demanding disciplines in advanced technology, generating buzz every when a novel product is launched by any company using ML techniques and algorithms for reaching consumers in extremely original ways. Machine learning is a method that enables computers to learn without being programmed. Despite people not being aware of it, machine learning is actively employed in daily life. Fundamentally, science is what makes it possible for machines to translate, process, and study data to address real-world issues. To create a complete ML system, programmers create machine learning algorithms that are written in machine language using complex mathematical expertise. Thus, with ML, we may categorise, understand, and estimate data from a given dataset. Data professionals choose different machine learning (algorithm) types depending on the sorts of data accessible for the predictions they wish to make from the data:

1. Supervised Learning: In this sort of learning, data professionals provide algorithms with labelled training data as well as defined variables for accessing and discovering correlations. The algorithm's input and output are specific and defined.
2. Unsupervised Learning: This kind of learning uses algorithms that study unlabeled data and examine datasets to identify relevant connections or relationships. To find hidden

patterns or groups in datasets, one way is cluster analysis, which makes use of exploratory data analysis.

3. Reinforcement Learning: Reinforcement learning is used to train a computer programme to complete a multi-step procedure with well-defined rules. Here, programmers create an algorithm to carry out a task and provide positive and negative feedback so that the algorithm will act as it should to carry out the task. In other cases, the algorithm even chooses for itself what to do next.

3.2.2 Deep Learning

Deep learning model is a type of AI and ML which learn like how people acquire specific types of information. Data science, which also encompasses statistics and predictive modelling, contains deep learning as a key component. Deep learning makes this process quicker and simpler, which is very advantageous to data scientists who are entrusted with gathering, analysing, and interpreting massive amounts of data. Deep learning may be conceptualised as an automated kind of predictive analytics at its most basic. Deep learning algorithms, in contrast to conventional machine learning algorithms, are layered in a hierarchy of increasing complexity and abstraction.

Self-learning representations are a property of deep learning algorithms, they also rely on ANNs that simulate how the brain processes information. In order to extract features, classify objects, and identify relevant data patterns, algorithms exploit unknown components in the input distribution throughout the training phase. This takes place on several levels, employing the algorithms to create the models, much like training machines to learn for themselves.Several algorithms are used by deep learning models. No network is seen to be flawless, although certain algorithms are better adapted to carry out particular tasks. It's beneficial to get a thorough grasp of all major algorithms in order to make the best choices. The following four stages are used to define and train all models:

1. Splitting dataset: The train test split typically consists of 80% training data and 20% test data. We use a 0.8 ratio for dividing the EEG dataset for all models in this study. A known output is included in the training set as a label, either one or zero, and the model gains knowledge from this data. to later be applied to data from other sources.

2. Forward Propagation: This phase involves determining activations at each layer. The ultimate output is produced as a result of these activations flowing forward from the input layer to the output layer.

3. Loss Computation: The loss or error is computed in the output layer in this stage. The variation between the actual result and the projected value may be determined using a straightforward error function.

4. Backpropagation: By making minor adjustments to the bias and the weights, this phase seeks to decrease error in the output layer. The error term's derivatives are used to calculate these changes. The delta changes are returned to hidden layers, where equivalent adjustments are performed to their weights and bias according on the Calculus concept of the Chain rule. As a result, the weights and bias are adjusted until the error is reduced to the absolute minimum.

5. Parameter updation: Finally, using the delta changes obtained from the previous backpropagation stage, the weights and bias are modified. The loss is decreased to a low value when these procedures are carried out for a number of epochs with a high number of training samples. On the basis of the unobserved data, predictions may then be made using the final weight and bias values.

3.3 Conclusion

The basic theory behind artificial intelligence,machine learning and deep learning are studied in this chapter. The next chapter deals with the different implemented deep learning models for the classification of alcoholic EEG data.

Chapter 4

IMPLEMENTED MODELS

4.1 Overview

This chapter deals with recently implemented deep learning methods to find better accurate classifier of alcoholic EEG signals for medical applications. By comparing the performance of proposed model's,it's possible to identify most accurate deep learning classifier for alcoholic EEG classification.

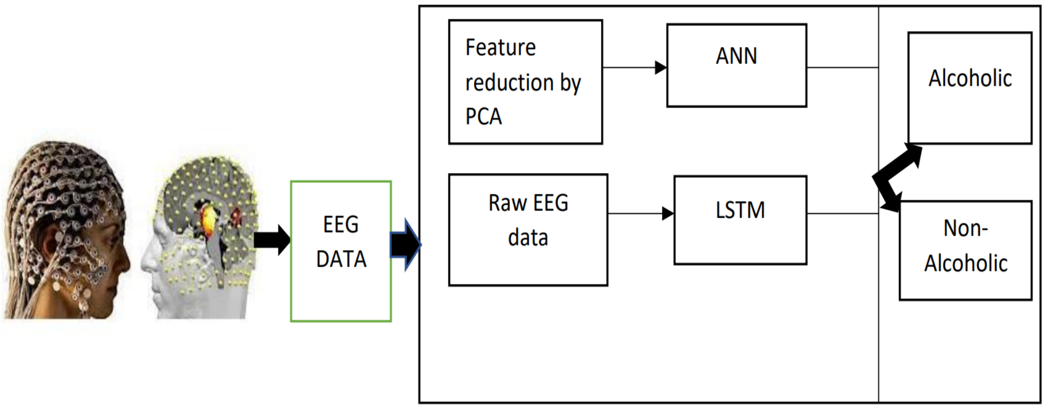

Figure 4.1: Block diagram of implemented methodology for classification of alcoholic EEG data.

4.2 Dataset

The current dataset is from UCI ,the EEG data is from the State University of New York's Neurodynamic Laboratory[17]. It comprises a total of 122 participants, including 77 people who have been diagnosed with alcoholism and 45 people who are not. A total of 120 trials were conducted for each individual. All trials were labelled as alcoholism if the subject was an alcoholic. With 64 electrodes put on the patient's scalps for 1 second, all trials were captured at 256Hz. The categorization procedure must determine if the person has been affected with alcoholism or whether they are a non-alcoholic group.

4.2.1 Data Acquisition Process

Electrodes are used to measure or record EEG signals, however electrode placement cannot be done at random; there are specific guidelines for that [20]. The 10-20 system, sometimes referred to as the International 10-20 system, is a method that is commonly used to describe and implement scalp electrode placement. The 10-20 system, which is shown in figure 4.2, is based on the relationship between an electrode's location and the region of the cerebral cortex underneath it. Each site is identified by a letter (to define the lobe) and a number or another letter (to denote the location of the hemisphere). The letters F, T, C, P, and O stand for frontal, temporal, central, parietal, and occipital. Even numbers (2,4,6,8) are associated with the right hemisphere, whereas odd numbers (1,3,5,7) are associated with the left. The z denotes the centre electrode.In this work 64 electrodes are placed on the scalp.

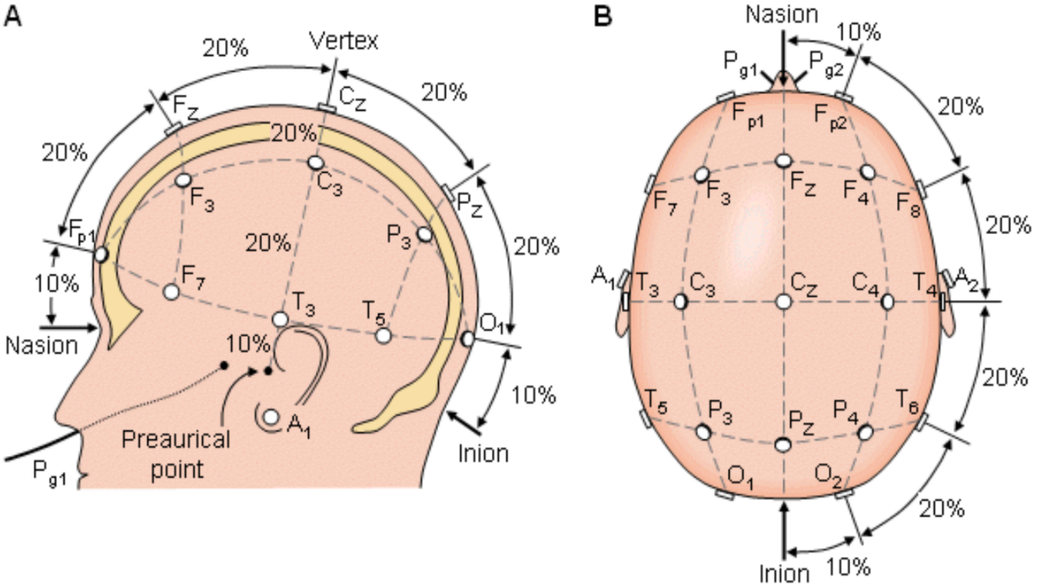

Figure 4.2: Electrode mapping-10-20 system

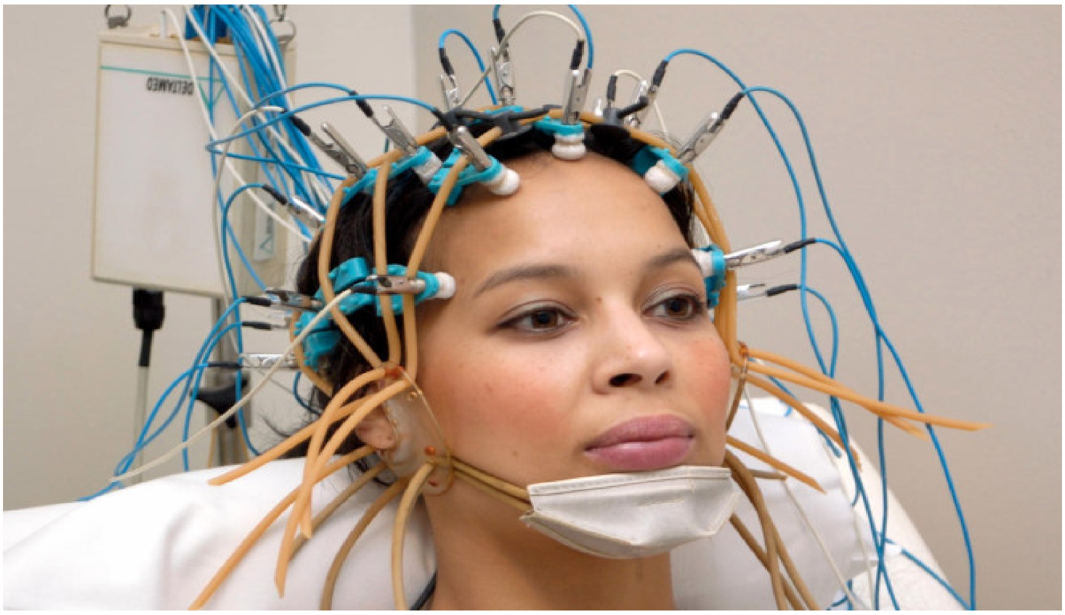

Figure 4.3: Electrode placement

4.3 Previous Methodology

Figure 4.1 shows the block diagram of previous methodology. It mainly focused on two deep learning models for classification of alcoholic and non-alcoholic EEG data.Model A is based on Artificial Neural Network(ANN) based classification and model B is based on LSTM method. In model A,raw EEG data is first applied to Principal Component Analysis(PCA) for feature extraction and reduction technique. Then the reduced components are given as input to the ANN model for classification of alcoholic EEG data[20]. In the model B raw EEG data is directly applied to LSTM for classification.

4.4 Dimension Reduction using PCA

Principal Component Analysis, or PCA, is a dimensionality-reduction method for lowering the dimensions of large data sets. It is done by downsizing a huge collection of variables while retaining the majority of the data in the larger data set. The primary components provide statistically important information about the real data, and they may be described as data variance or a technique for identifying the components that cover the largest variation. They can also be used to reproduce the real data set. The PCA approach is used for this work because EEG data is always complicated and is high-dimensional, making it difficult to employ as an input to a classification model directly.Reduced dimension is intended to cause the least amount of data loss possible[18] . After PCA, the majority of the data are in a reduced dimensional space, which makes them ideal to use as an input to a deep learning or machine learning classifier. It means,since EEG data involves informations from numerous sites on the human scalp and the collected data contains mixed brain activity, this approach is used to compute the independent components in order to see the actual properties of neuron activity. Increasing the speed of the classifier is one of the most significant uses of the PCA method. The orginal dataset have 64 columns which means it have 64 features. By applying PCA method we reduces the number of features from 64 to 30 for applying to ANN. These 30 components are considered as relevant features for reflecting the properties of the actual EEG data.

4.5 Mathematical working of PCA

1. Take features of the dataset without considering Label.

2. Normalisation of Dataset.

3. Calculate Covariance Matrix.

$$\mathrm{cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(Xi - \vec{x})(Yi - \bar{y}) \tag{4.1}$$

4. Calculate Eigen Values.

$$A \cdot v = \lambda \cdot v \tag{4.2}$$

5. Calculate Eigenvectors.

$$(A - \lambda \cdot I) \cdot v = 0 \tag{4.3}$$

   Let A be a square matrix, v a vector and λ a scalar that satisfies equation 4.2, λ is eigenvalue associated with eigenvector of A.

6. Sort the eigenvectors by corresponding to decreasing eigenvalues and choose first n eigenvectors with the largest eigenvalues.

7. Create final data points with an eigenvectors matrix.

$$\text{(Normalised A matrix) . (Eigenvector Matrix) = Transformed Dataset} \tag{4.4}$$

4.6 Classification Using ANN

Artificial neural network (ANN) is a system inspired by the brain, having three layers: an input layer, an intermediate layer, and an output layer.The inputs from the outside world is passed to the input layer.The hidden layer or intermediate layer consist of a set of neurons and all computational process is occuring here.The intermediate layer's nodes have the ability to convert the input into anything the output layer may utilise, allowing the layers extract various attributes until the network understands what they is searching for. The ANN training procedure is comprised of comparatively quicker than other varieties of deep learning techniques.One of the most typical time series data types is EEG signals, which problems with the ANN network include excessive training and responsiveness to the quantity of hidden neurons.

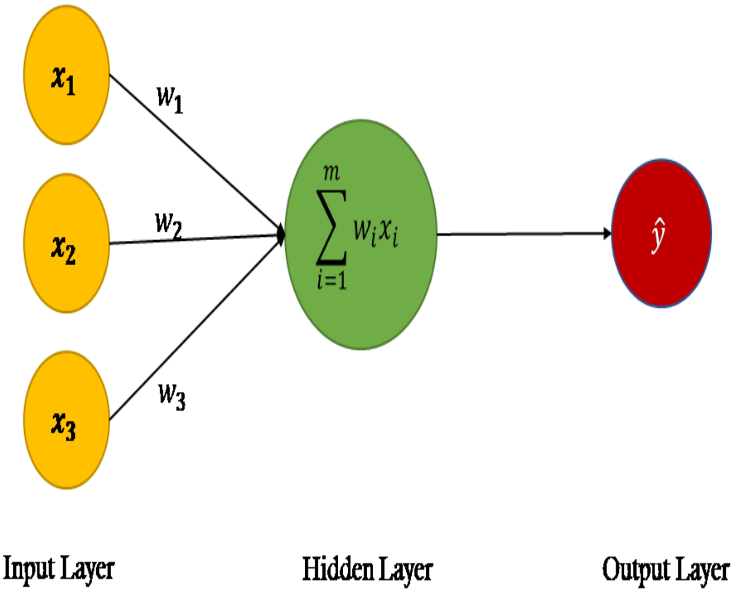

Figure 4.4: Structure of ANN.

ANNs are typically categorization techniques made up of a large number of computationally simple linked neurons. The number of nodes in the first layer depends on the dimensions used. Here, employed 64 channels from the raw EEG dataset and 30 main components from the PCA method's output for working with both ANN and LSTM.

Figure 4.4 shows the basic structure of ANN.Here inputs(X) are multiplied with their assigned weights(W). After assigning weights, bias variable(b) is added.Bias is a constant that helps the model to fit in the best way possible. Next is to apply activation function to linear equation and finally moves to output layer.

$$z_1 = X^* \, W_1 + b_1 \tag{4.5}$$

The number of desired classes determines how many neurons are present in the output layer. In this study, we need to establish if a subject is an alcoholic or not, hence number of class is 1. It is essential to add intermediate or hidden layers to boost the network's capability. There are no restrictions on how many layers and nodes are required. Smaller numbers of layers and nodes result in mistakes, poor performance, and poor generalisation, whereas large numbers of hidden layers and neurons make the network more complicated and lengthen the execution time. Since there are no guidelines to establish the optimum topology, it can only be discovered by trial and error.

4.7 Design of PCA and ANN

The proposed approaches are tested with UCI alcoholic EEG data . 64 channels of data from 122 participants make up the UCI EEG data,out of this, 77 of total subjects were diagnosed with alcoholism and 45 of them were not. In this section , for the classification of EEG data, ANN is employed as a classifier in two different experiments named as model 1 and model 2.

Model 1

The model combines the PCA output with a straightforward multi-layer ANN. The 50 neurons in the intermediate layer are used to learn from the input, and the nodes use a tangent activation function. The input neurons in the first layer consist of 30 major components. The output layer has one node for each potential type of alcoholic or not, and it is the last layer. This categorization was carried out using a sigmoid output layer.

Model 2

The four-layer ANN model is fed with raw EEG data. Three intermediate layers with a total of 100, 50, or 32 neurons each are used to learn from the input; the nodes use a Relu activation function. The first layer has 64 channels as input neurons. The only category of alcoholic or not corresponded to an output layer with a single node. This categorization was carried out using a sigmoid output layer. The key benefit of utilising a sigmoid function for the last layer is that its output may range from zero to one, making it helpful for models in which the probability must be predicted as the outcome. ReLU functions perform better overall in our ANN models for the intermediate layers.

4.8 Long Short-Term Memory (LSTM)

A multi-layer neural network called a recurrent neural network (RNN) is used to forecast sequential input, such voice recognition. The approach makes use of weighted memory and a feedback system.Standard RNN suffers from short term memory due to vanishing gradient problem then form when working with large data.When dealing with timeseries data, such as EEG signals, the Long Short-Term Memory (LSTM) approach, a kind of RNN, is particularly effective. The Long Short-Term Memory (LSTM) model was used here because it can store crucial data about earlier states and take advantage of the time dependencies between the data utilising a memory cell.

This network is appropriate for EEG applications because of LSTM's memory for past observations. Information moves from one layer to another layer in a conventional ANN. Since the layers have no recollection of prior states, they are stateless. RNNs enable loops by enabling output nodes to act as input nodes. Although RNNs are exceedingly challenging to train and frequently overfit very quickly, this makes them particularly effective networks for time series issues. The LSTM approach, a form of recurrent neural network, was developed in 1997 by Hochreiter and Schmidhuber to address the issues of long-term reliance and gradient explosion in recurrent networks[19] .When a significant mistake builds up during the training process, it can cause very large updates, which is known as exploding or disappearing gradients.Figure 4.5 shows the basic structure of LSTM. LSTM consist of three gates:forget gate,input gate, output gate.Forget gate controls what information should be forgotten. Sigmoid function ranges between 0 and 1, it sets which values in the cell state should be discarded or remembered and tanh function ranges between -1 and 1 which helps to regulate the values flowing through the network. Input gate helps to identify important elements that need to be added to the cell state.

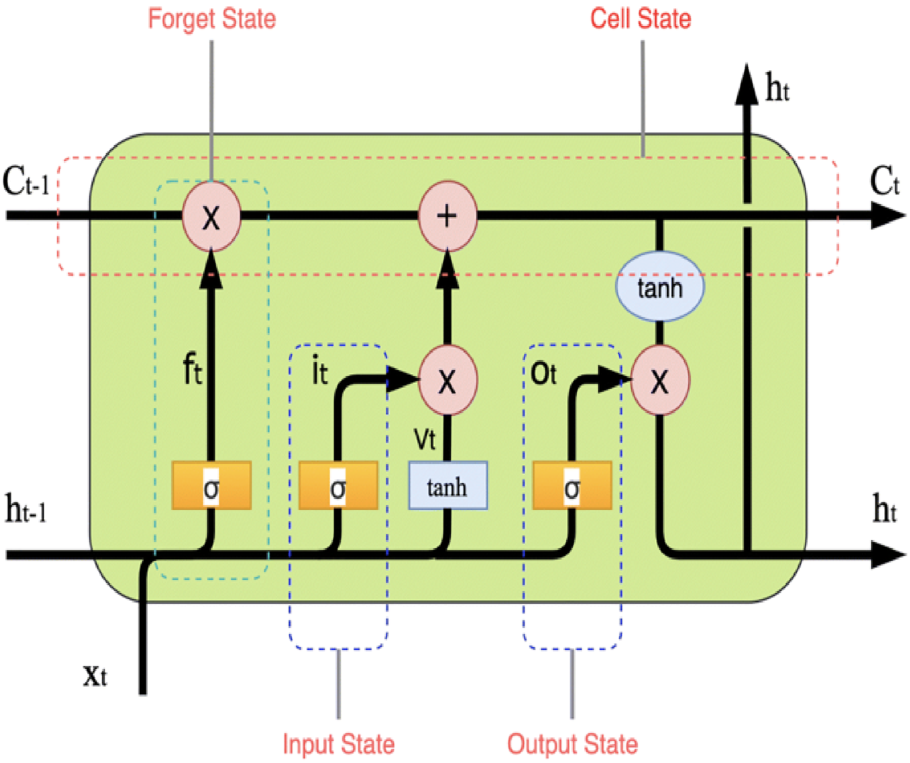

Figure 4.5: Structure of LSTM

4.8.1 Forget gate

$$f_t = \sigma_g \left(W_f \times x_t + U_f \times h_{t-1} + b_f \right) \tag{4.6}$$

t=time step

f_t= forget gate at t

x_t=present input

h_{t-1}= previous hidden state

b_f=connection bias at t

The forget gate (eq.4.6) decides which information needs to be added and which can be ignored. The present input $x(t)$ and hidden state $h(t-1)$ are passed through the sigmoid function. Sigmoid function regulates the values between 0 and 1. It concludes whether the part of the old output is necessary or not. This value of $f(t)$ will later be used by the cell for point-by-point multiplication with previous cell state.

4.8.2 Input Gate

$$i_t = \sigma_g \left(W_i \times x_t + U_i \times h_{t-1} + b_i \right) \tag{4.7}$$

$$v_t = \tanh_c \left(W_c \times x_t + U_c \times h_{t-1} + b_c\right) \tag{4.8}$$

t=time step

i_t = input gate at t

v_t = value generated by tanh function

x_t =present input

h_{t-1} = previous hidden state

b_i,b_c= bias vectors

The next step is to decide what datas are going to store in cell state.). First, the current state x_t and previously hidden state $ht-1$ are passed into the second sigmoid function. The values are transformed between 0(forget) and 1 (keep value in network). Next, the same information of the hidden state and present state will be passed through the tanh function for regulating the network.The output of tanh is v_t with all the possible values between -1 and 1.The output values generated from this step is then point-by-point multiplicated.

4.8.3 Cell State

$$c_t = f_t \cdot c_{t-1} + i_t \cdot v_t \tag{4.9}$$

where,t= time step

c_t = cell state information

f_t = forget gate at t

c_{t-1} = previous cell state at $t-1$

The input gate and forget gate have provided the network with sufficient data. Next step is to update old cell state to new cell state c_t. The forgets vector f_t is multiplied by the prior cell state c_{t-1} . The network takes the output value of the input vector i_t and performs point-by-point addition, introducing a new cell state c_t to the network.

4.8.4 Output Gate

$$o_t = \sigma_g \left(W_o \times x_t + U_o \times h_{t-1} + b_o\right) \tag{4.10}$$

$$h_t = o_t \cdot \tanh_c \left(c_t\right) \tag{4.11}$$

The LSTM has an input $x(t)$. (h_t-1) and $(c_t-1))$ are the inputs from the previous timestep LSTM. $o(t)$ is the output of the LSTM for this timestep in equation 4.10 . The LSTM also generates the $c(t)$ and new hidden state $h(t)$ for the consumption of the next time step LSTM.

4.8.5 Implementation of LSTM

Here,122 subject's raw EEG signals are applied to the LSTM deep learning algorithm to determine whether or not they are alcoholics. Each participant is equipped with a data matrix and labels. Each subject's class is represented in label vector by one (non-alcoholic) or zero (alcoholic).One dropout layer, one dense layer, and two fully linked LSTM layers make up the LSTM model. Probability is the foundation of dropout works. They are built by deleting inputs to a layer, which might be activation nodes from a previous layer or input characteristics in the first layer from the raw data set. This has the effect of shrinking a huge network to a distinct, smaller network structure, which strengthens the network's neurons as a result. 64 characteristics are extracted from the raw EEG signals using the LSTM and dropout layers, and the dense layer is then utilised for classification.

The LSTM model is trained on 80 percent of the data using 10-50 epoch iterations and tested on 20 percent of the EEG data.

4.9 Conclusion

This chapter addressed the different deep learning models used for classification of alcoholic EEG signals .Next chapter deals with the proposed models to classify EEG data's to overcome the limitation in performance of past implemented models.

Chapter 5

Proposed Methodology

5.1 Overview

In this chapter to overcome the limitation in accuracy of ANN and LSTM models outlined in previous section, proposing a comparison of some deep learning, one machine learning and some Ensembled models for proposing the most accurate classifier for alcoholic EEG signals.

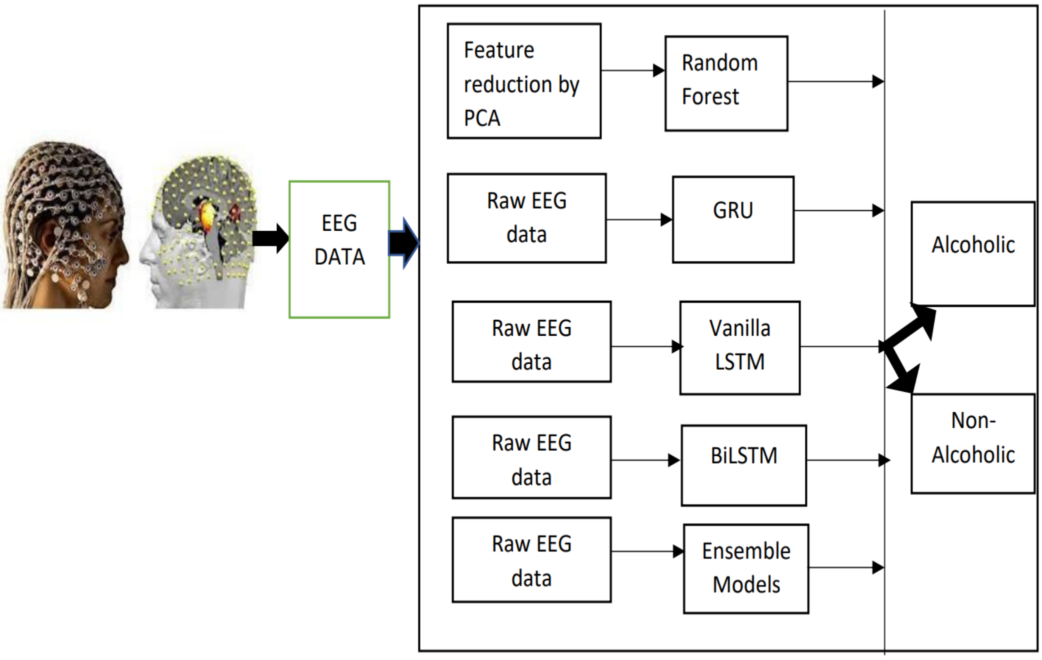

Figure 5.1: Block diagram of proposed methodology

Algorithm 1 is focused on Random Forest model.Algorithm 2,Algorithm 3 and Algorithm 4 is based on deep learning models such as Gated Recurrent Unit(GRU),Vanilla Long Short Term Memory and Bidirectional Long Short Term Memory (BiLSTM) respectively and Algorithm 5 is a ensembled model. In Algorithm 1,the raw EEG data is initially subjected to Principal Component Analysis (PCA) in order to find significant components and minimise the feature dimension. The reduced features are then fed into Random Forest model that separates intoxicated data and control data.In Algorithm 2,Algorithm 3 and Algorithm 4, raw EEG data are directly fed to deep learning models.Finally in ensembled model, different classifiers are combined for classification. The proposed approaches are described below, along with the implementation procedure.Block diagram of proposed work is shown in figure 5.1.

5.2 Random Forest

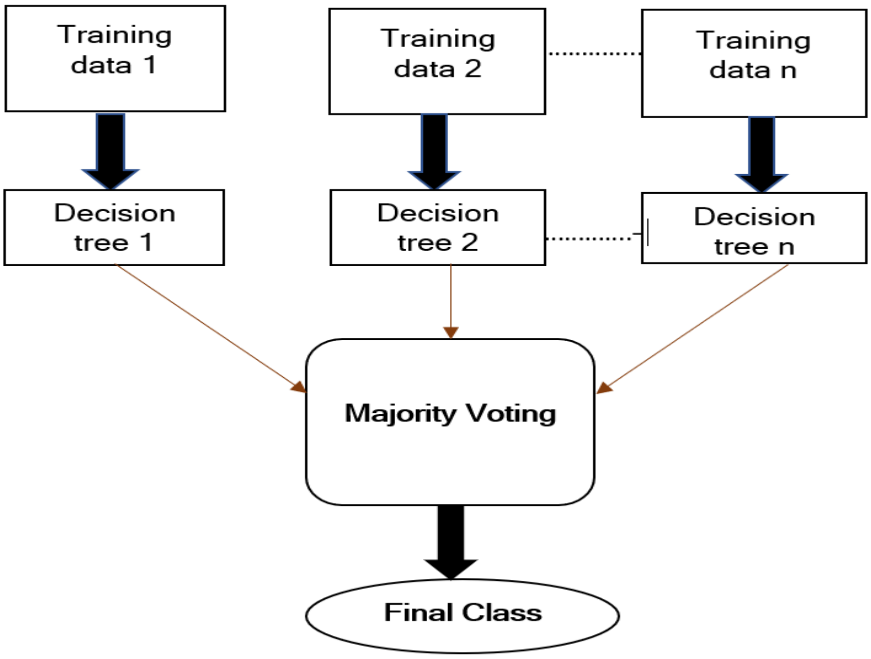

Figure 5.2: basic structure of random forest

Random forest is a Supervised Machine Learning model which consist of many decisions tree that is used widely in both Classification and Regression problems. Random Forest is one of the most important algorithm in machine learning. The Random Forest Algorithm's ability to handle data sets with both continuous and categorical variables, as in regression and classification, is one of its key features.Each decision tree consist of three nodes which are root node,decision node and leaf node.

Here the entire data set is fed at root node.At decision node decisions are taken and splitting is performed. Leaf node is the ending point of tree where nofurther division takes place.

One of the main limitation of random forest is that large number of trees can make the model too slow. Random forest works based on bagging principle. In random forest, bagging choose some samples from the whole dataset. Individual decision trees are made for each samples. Final output is based on majority voting of each output generated by decision trees.

5.2.1 Classification using Random Forest

In two separate experiments, this study used 64 channels from the raw EEG dataset and 20 principal components from the PCA method's output. Higher accuracy comes at the cost of slower learning but when there are more decision trees. More accuracy means more execution time. Here trial and error is performed to determine the best topology.The proposed approaches are tested with UCI alcoholic EEG data . 64 channels of data from 122 participants make up the UCI EEG data,out of this, 77 of total subjects were diagnosed with alcoholism and 45 of them were not.In this section of study for the classification of EEG data, Random Forest is employed as a classifier in two different experiments named as model 1 and model 2.

Model 1

The model 1 is a combination of output of PCA plus Random Forest model. Here 20 principal components are given as input to the root nodes.Gini impurity is used as the splitting criterion here,and the number of n estimators is varied between 10 and 30 for better accuracy analysis.

Model 2

In model 2 raw EEG data is directly applied to Random Forest model. Gini impurity is used as the splitting criterion here,and the number of n estimators is varied between 10 and 30. The final output will be one class,either alcoholic or non-alcoholic based on majority voting.

5.3 XG Boost

XG Boost means Extreme Gradient Boosting. Decision trees are built sequentially in XG boost and weights are significant in XGBoost. Each independent variable is given a weight before being put into the decision tree that forecasts outcomes. Variables that the tree incorrectly anticipated are given more weight before being placed into the second decision tree. These distinct classifiers/predictors are then combined to produce a robust and accurate model. It may be used to solve issues including regression, classification, ranking, and custom prediction. XG boost is used in the ensembled model in this work.

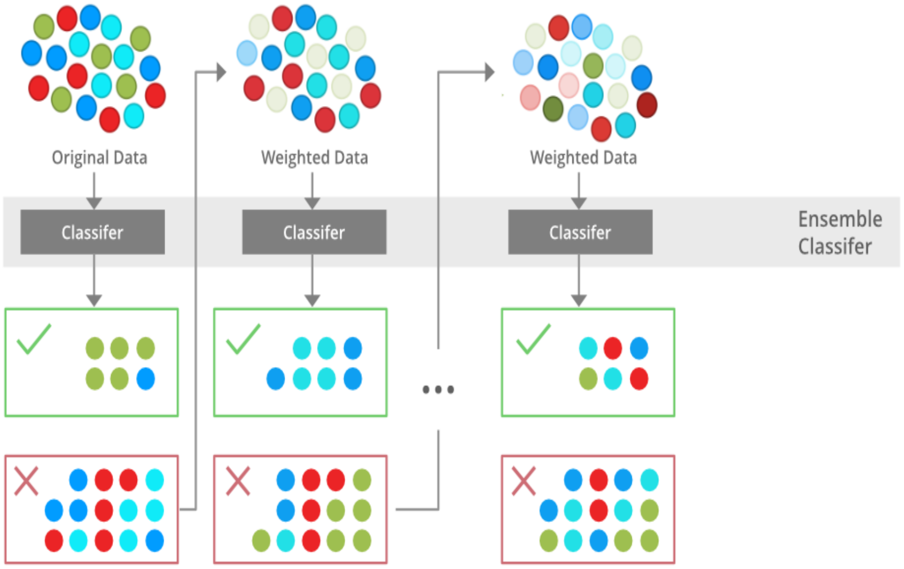

Figure 5.3: basic structure of XG-boost

5.4 Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a updated version of recurrent neural network that looks a lot like an LSTM.GRU were created to avoid vanishing gradient problem which comes with a standard recurrent neural network(RNN). GRU can considered as a variation on the LSTM because both are designed similarly and in some cases, both produce equally good results. The GRU's abandoned the cell state in favour of using the hidden state to transfer data. There are just two gates on it: a reset gate and an update gate.The update gate, functions similarly to an LSTM's forget and input gates. It chooses which information to discard and which information to include. Another gate used to select how much past information to forget is the reset gate. GRU is quicker and consumes less memory than LSTM,although it's not possible to say which model produce better accuracy,because it also depends on the type of dataset. Here evaluated different GRU designs by varying different parameters on a sample of data like varying acti-

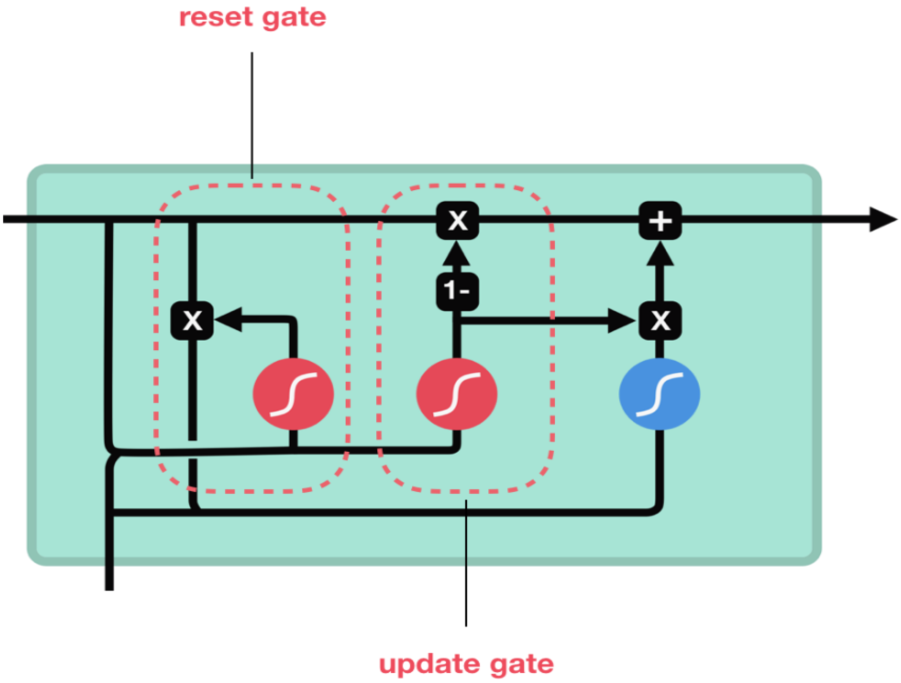

Figure 5.4: basic structure of gru

vation function, dropout value, batch size and by adjusting the number of layers, neurons per layer.

5.4.1 Implementation of GRU based model

The GRU algorithm is given to raw EEG data of 122 subjects to determine alcoholic or not . Each participant has a matrix of data and labels. Each subject's class is represented by zero (alcoholic) or one(non-alcoholic) label vector. Dropout is based on probability.The input layer, GRU layer, a dropout layer with probability 0.4, and a dense layer for classification comprise the structure of the GRU network. The GRU model is trained on 80% of the total data and tested on 20% of the EEG data using 10-50 epoch iterations.ADAM optimizer were used for this design.

5.5 Vanilla LSTM

An LSTM model with a single hidden layer of LSTM units and an output layer is known as a Vanilla LSTM. The model is fit using the efficient Adam optimizer and optimized using the mean squared error, or 'mse' loss function. Using 10 to 50 epoch iterations, the Vanilla LSTM model is trained on 80% of the data and tested on 20% of the total EEG data.The input layer, a LSTM layer, and a dense layer for binary classification comprise the structure of the vanilla LSTM network.

5.6 Bidirectional LSTM

A bidirectional LSTM(BiLSTM), often known as a biLSTM, is a two-LSTM sequence processing model, one of which receives input in the forward direction and the other in the backward direction. BiLSTM's effectively enhance the quantity of data available to the network, giving the algorithm better context.

Recurrent Neural Networks (RNNs) are multi-layer neural networks that are used to estimate sequential input, such as voice recognition. The technique makes use of a feedback loop and weighted memory. Exploding or vanishing gradients are a major problem with RNN. Due to this issue ,during back propagation time ,the weights won't get updated.New weight will be equal to old weight in this case. When dealing with timeseries data such as EEG signals,Long Short-Term Memory (LSTM) and Bi-directional Long Short-Term Memory (BiLSTM) are particularly effective. Unlike standard LSTM,in BiLSTM the input flows in both directions, and it's capable of utilizing information from both sides. A bidirectional LSTM differs from a conventional LSTM in that our input flows in two directions. It is a varient of LSTM that helps to

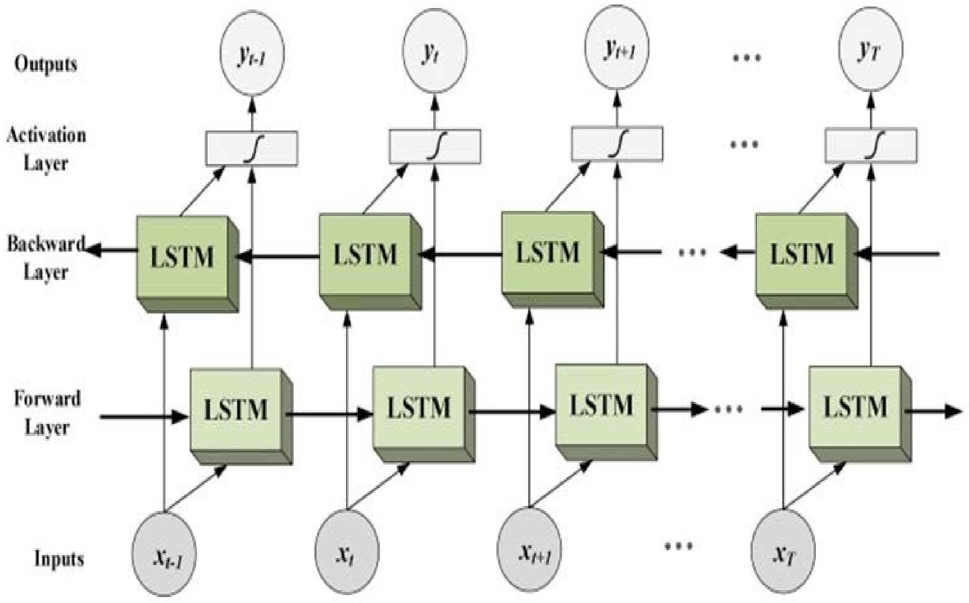

Figure 5.5: basic structure of bilstm

improve model performance. We may make input flow in one way, either backwards or forward, using the standard LSTM. However, with bi-directional input, we can maintain both future and historical information by allowing input to flow in both directions. Bi-directional model allows us to preserve both past and future information by allowing input to flow in both directions which makes this model suitable for EEG data classification.

The information flow is reversed with the addition of one more LSTM layer by BiLSTM. It simply suggests that the input sequence flows backward in the additional LSTM layer. The outputs from the two LSTM layers are then merged in a variety of methods, such as by averaging, adding, multiplying, and concatenating them.

5.6.1 Proposed BiLSTM Model Design

The BiLSTM algorithm is given to raw EEG data of 122 subjects to determine alcoholic or not.A fully connected BiLSTM layer, one dropout layer, and one dense layer make up the model.The BiLSTM model is trained on 80% of the total data and tested on 20% of the EEG data using 10-50 epoch iterations.ADAM optimizer is used here for reducing the model loss and to increase the accuracy.

5.7 Ensemble Methods

Ensemble models is formed by mixing numerous models rather of relying just on one, ensemble approaches seek to increase the accuracy of results in models[22]. The integrated models considerably improve the accuracy. For both regression and classification, ensemble approaches work's best because they lower bias and variance and increase model accuracy.The models used for ensemble are called Base estimators. It provides a way around the following technical difficulties in creating a single estimator:

1. High variance
2. Low accuracy
3. Features noise and bias

Figure 5.6 shows the basic structure of a ensemble classifier model. The inputs are distributed to each classifiers in an ensemble learning architecture. The output of these classifiers are

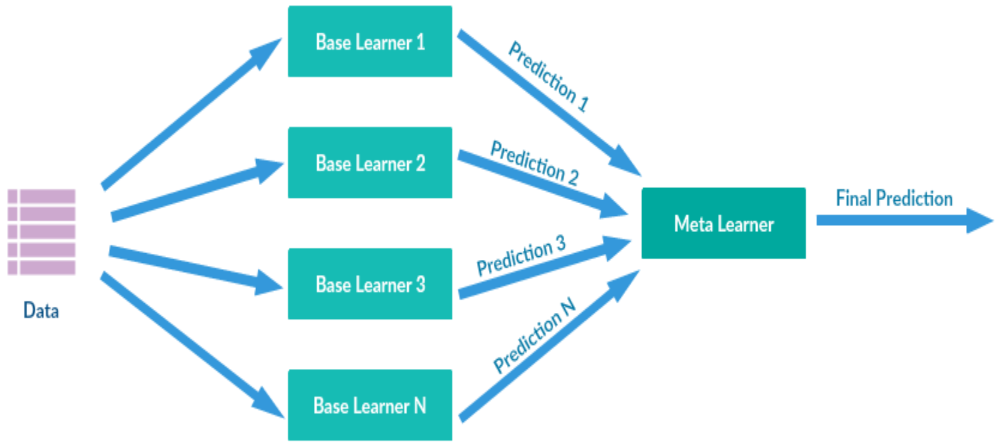

Figure 5.6: Basic Ensemble Technique

then given to a voting classifier. A voting classifier is a machine learning model that gains information by training on a collection of several models and predicts an output (class) based on the class with the highest probability of being the output. To forecast the output class based on the largest majority of votes, it simply sums the results of each classifier that was provided into the voting classifier. In this work we use hard voting. In a hard voting, the output class that is predicted is the one that received the greatest number of votes, or the class that had the greatest chance of being predicted by all of the classifiers. If three classifiers predicted the output class (A, A, and B), the majority would have projected A as the output in this case. As a result, the outcome will be A.

Here random forest and XG-boost are used as the base estimators. The output of the base estimators is given to a voting classifier. Voting classifier or meta learner used in this study is random forest algorithm. And the final output of the voting classifier is one with majority votes.

5.8 Conclusion

The proposed models required for this work ,for classification of alcoholic and non-alcoholic EEG signals is discussed in this chapter. Next chapter deals with obtained results of all model.

Chapter 6

Experimental Results And Discussions

6.1 Overview

In this chapter, the accuracies of the different DL models and their performance are verified. As previously indicated, we used an EEG dataset from UCI for our investigation, which contains EEG signals from 122 patients. 2,804,224 vectors of 64 dimensions of the UCI EEG data where used for each method. Out of this, 2,243,379 vectors are employed as training set in each technique,and 560,845 vectors of same dimensions are served as the testing dataset. The training set is used to train all models, and the testing data set is used to evaluate performance for various setting for hyper parameters.On the UCI Alcoholic EEG dataset, the suggested algorithm's performances were evaluated in terms of accuracy,recall,precision and f1-score,obtained from confusion matrix.

6.2 Confusion Matrix

A table called a confusion matrix is used to describe how well a classification system performs. A confusion matrix shows and sums up a classification algorithm's performance.

Parameters of confusion matrix

1. True Positive (TP):The actual value and predicted value are same.The model predicts a positive value, and the actual result is positive here.

2. True Negative (TN):Here actual and predicted values are same.Both predicts a negative value.

3. False Positive (FP):It's a type of error.The model predicts a positive result, but the actual value was negative.

4. False Negative (FN) :A type of error.The model predicts a negative result, but the actual value was positive.

Accuracy,Recall,Precision and F1-score can be calculated in terms of above parameters.

$$\text{Accuracy } = \frac{TP + TN}{TP + FP + TN + FN} \tag{6.1}$$

$$\text{Precision } = \frac{TP}{TP + FP} \tag{6.2}$$

$$\text{Recall } = \frac{TP}{TP + FN} \tag{6.3}$$

$$\text{F1 score } = 2 * \frac{\text{Precision } * \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{6.4}$$

6.3 Grouping Performance Parameters

By evaluating the classification accuracy, recall, precision, and F1 score, we evaluated the effectiveness of the suggested methods and modified the hyperparameters of each model as necessary. Examples of these hyper-parameters include hidden layers, activation function, number of neurons, learning rate, number of epochs for training, batch size, and probability of dropout.

6.4 Vary Activation Function

Here we tried with both ReLu and sigmoid functions.ReLU's gradients do not become close to zero when x is very large, unlike sigmoid.ReLU converges more quickly than other activation functions.Sigmoid function is mainly used in output layer to perform binary classification.

6.5 Changing number of Neurons

The count of neurons is crucial because if there aren't enough nodes in each layer, the network won't be able to simulate complicated data, and the fit that results would be unsatisfactory.

If many neurons are employed, the network may overfit the dataset, which would result in an overly extended training period.Overfitting can generate random disturbance in the EEG data.This helps the model to fit with training data with good accuracy but it won't be able to generalize new unseen data.

6.6 Experimental Results of Implemented Models

Previous Studies showed that significant improvements have been gained with the LSTM method (Model B) (92.74% accuracy) and shows that LSTM is a robust and reliable classifier for EEG signals. The result for the ANN approach (Model A) is clearly below (91.60% accuracy) than the results for the LSTM approaches (Model B), however, the LSTM was not quicker than ANN and its run time was high.

6.6.1 Performance of ANN with PCA

Model Accuracy Plot

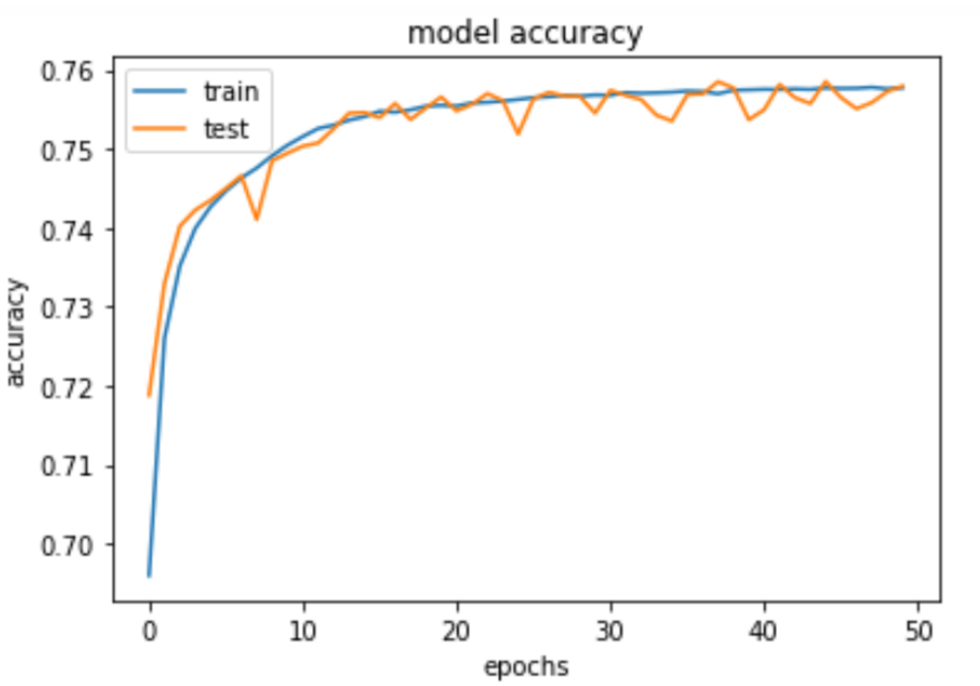

Figure 6.1: Accuracy Graph of ANN with PCA

Fig 6.1 shows the accuracy graph of ANN model with PCA.From graph it's clear that the accuracy increases from 70% to 75% at the end of iteration.

Model Loss Plot

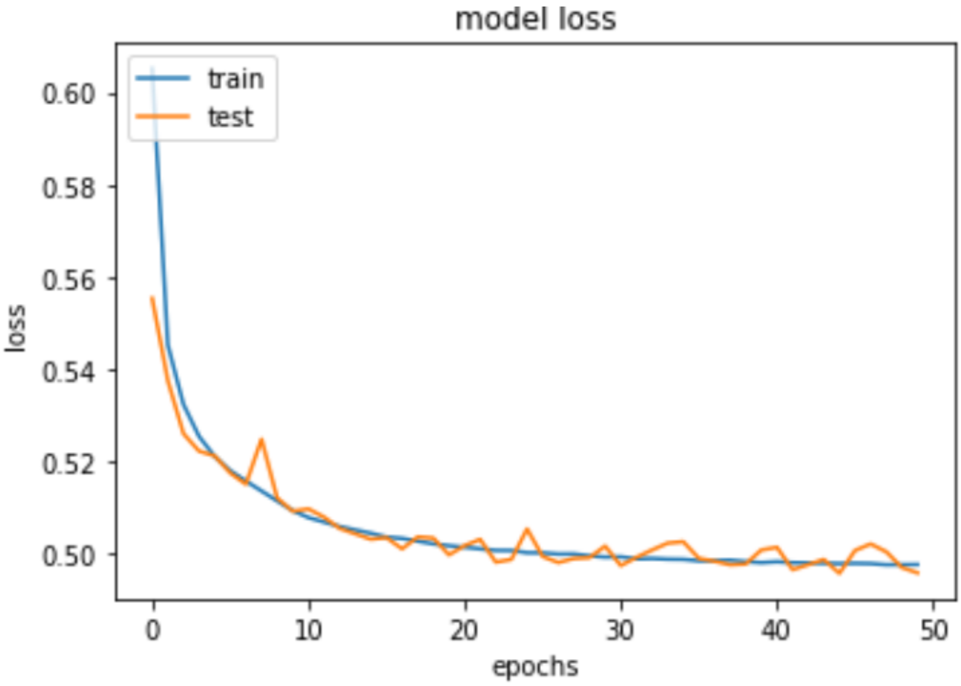

Figure 6.2: Loss Graph of ANN with PCA

From figure 6.2 it's clear that the loss function get reduced from 0.6 to 0.5 at the end of epoch in case of ANN with PCA.

Confusion Matrix

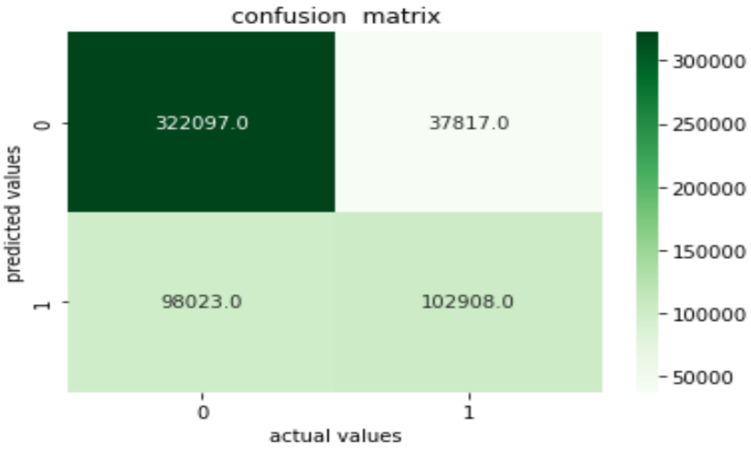

Figure 6.3: Confusion matrix of ANN-PCA model

Results from Confusion Matrix

Table 6.1: Values from confusion matrix of ANN-PCA model

TP	TN	FP	FN
322097	102908	37817	98023

Table 6.2: Performance parameters of ANN-PCA model

Accuracy (%)	Precision	Recall	F1 score
75.77	0.89	0.76	0.825

6.6.2 Performance of ANN without PCA

EEG data is directly applied to ANN model.

Model Accuracy Plot

Fig 6.4 shows the accuracy graph of ANN model.From graph it's clear that the accuracy increases from 75% to 91.6% at the end of iteration.

Model Loss Plot of ANN without PCA

From figure 6.5 it's clear that the loss function get reduced from 0.5 to 0.2 at the end of epoch in case of simple ANN model.However ANN without PCA has a better accuracy compared to ANN with PCA,but it takes more execution time.

Confusion Matrix

Figure 6.6 represents the confusion matrix of ANN model.

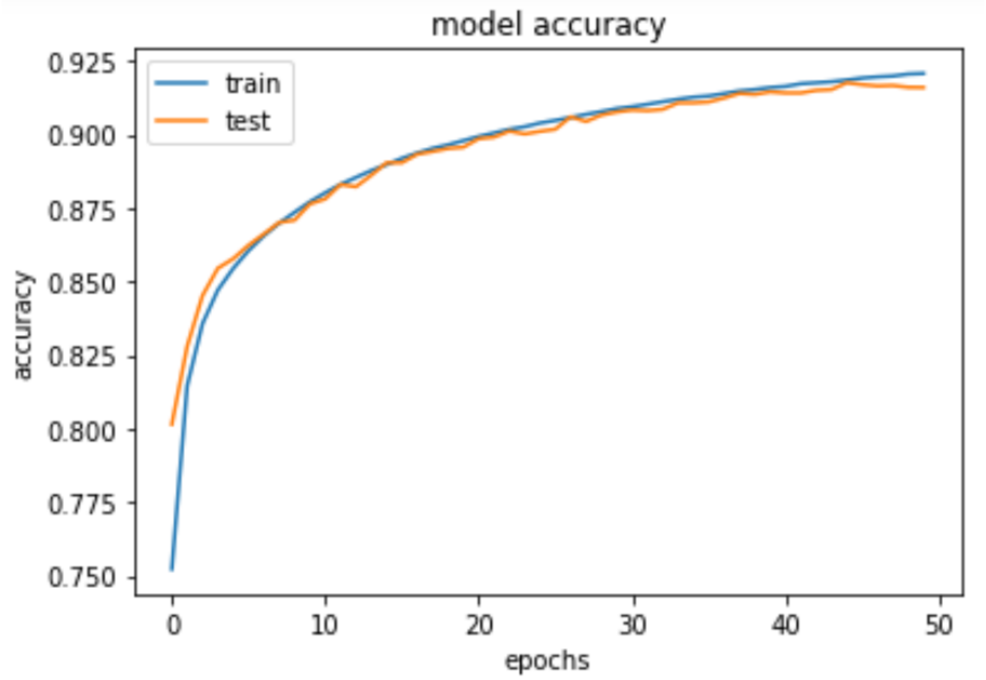

Figure 6.4: Accuracy Graph of ANN without PCA

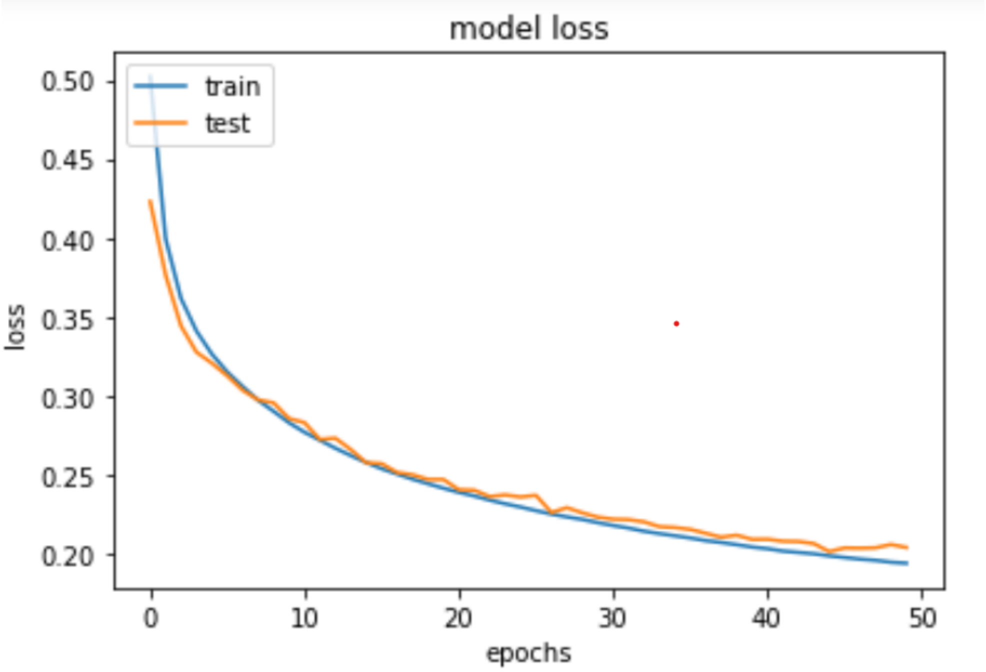

Figure 6.5: Loss Graph of ANN

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.3,and the performance parameters for evaluating the model is represented in table 6.4.

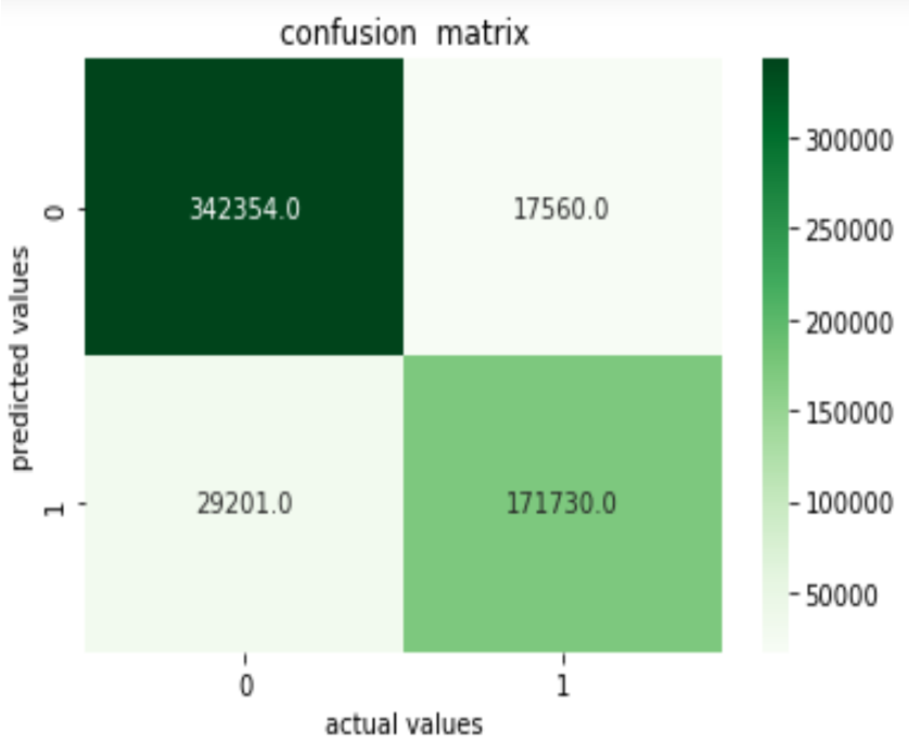

Figure 6.6: Confusion matrix of ANN model

Table 6.3: Values from confusion matrix of ANN model

TP	TN	FP	FN
342354	171730	17560	29201

Table 6.4: Performance parameters of ANN model

Accuracy (%)	Precision	Recall	F1 score
91.6	0.95	0.92	0.93

6.6.3 Performance of LSTM model

Model Accuracy Plot

Figure 6.7 shows the accuracy epoch plot of LSTM model.The accuracy gets increased from 80% to 92.7% at the end of iteration.

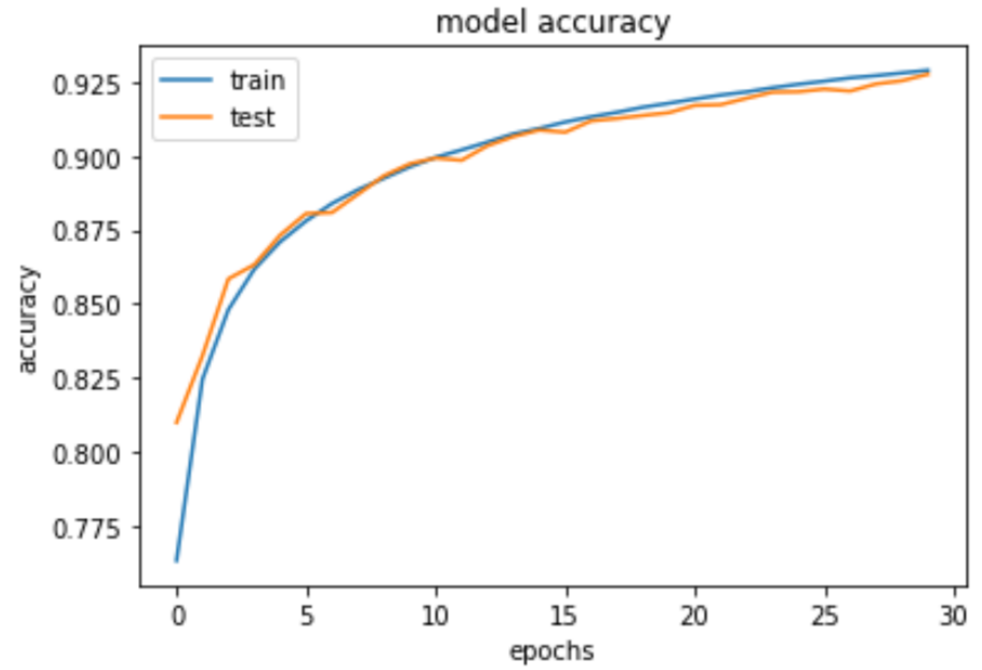

Figure 6.7: Accuracy Graph of LSTm model

Model Loss Plot

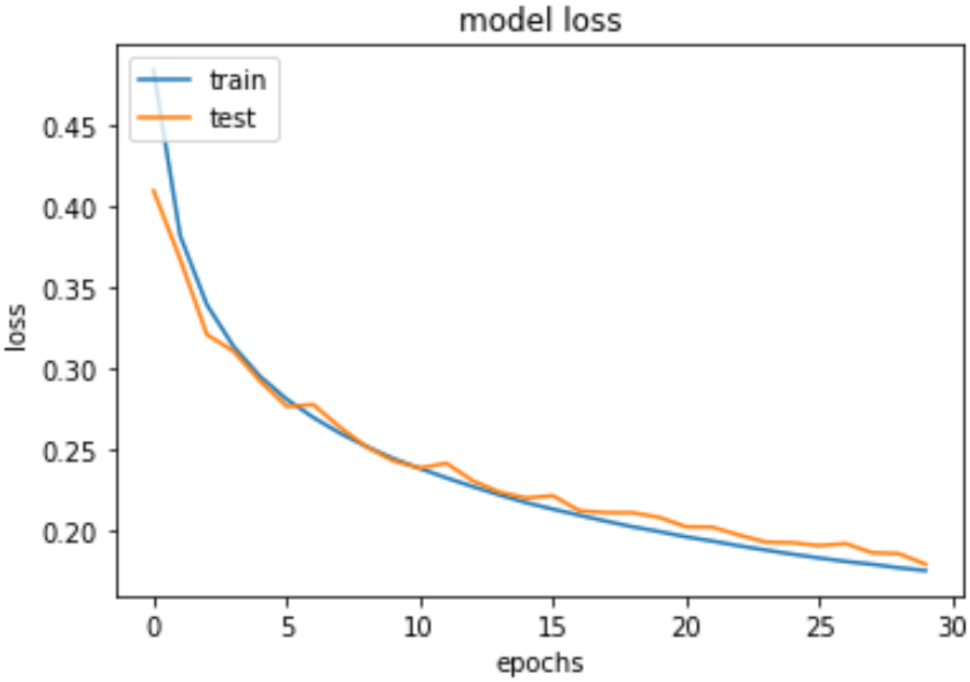

Figure 6.8: Loss Graph of LSTM model

Figure 6.8 shows the loss-epoch plot of LSTM model.The loss function started from 0.45 and at the end of epochs it decreased to 0.2

Confusion Matrix

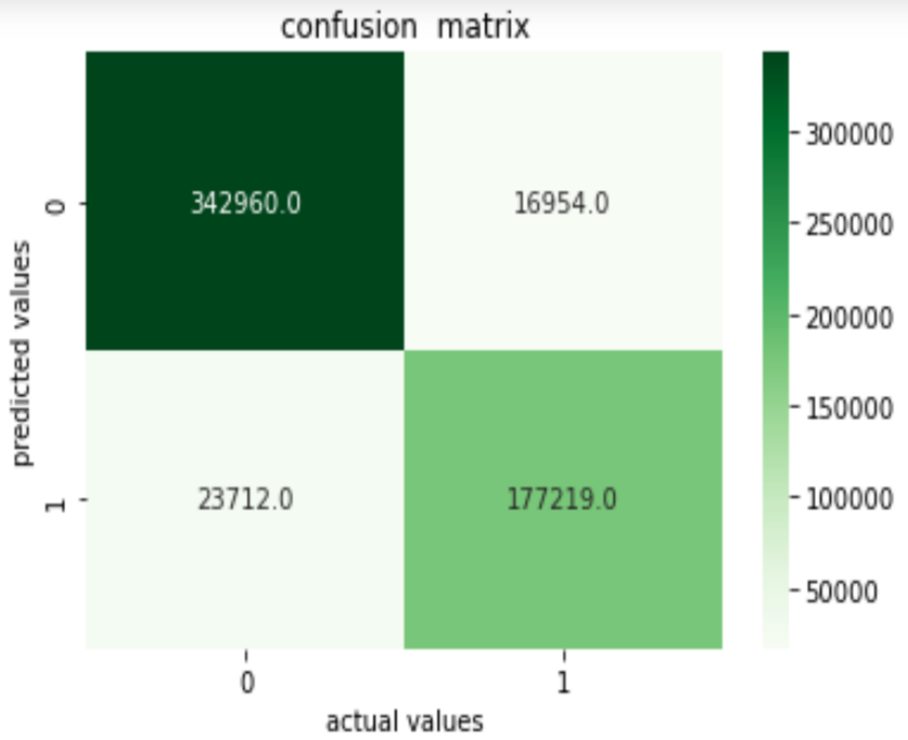

Figure 6.9: Confusion matrix of LSTM model

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.5,and the performance parameters for evaluating the model is represented in table 6.6.

Table 6.5: Values from confusion matrix of LSTM model

TP	TN	FP	FN
342960	177219	16954	23712

Table 6.6: Performance parameters of LSTM model

Accuracy (%)	Precision	Recall	F1 score
92.74	0.952	0.935	0.943

6.6.4 Performance of Implemented deep learning Models

Table 6.7: Comparison of implemented deep learning models

Model	Accuracy (%)	Execution Time	Precision (%)	Recall (%)	F1-Score (%)
PCA+ANN	75.77	02:50:00	89.4	76	82.5
ANN	91.6	06:00:00	95.12	92.14	93
LSTM	92.74	08:45:00	95.28	93.533	94.39

6.7 Experimental Results of Proposed Model

Here ,used an EEG dataset from UCI for our investigation, which contains EEG signals from 122 patients. 2,804,224 vectors of 64 dimensions of the UCI EEG data where used for each method.Out of this, 2,243,379 vectors are employed as training set in each technique,and 560,845 vectors of same dimensions are served as the testing dataset. Four experiments are run on this dataset using some machine learning and deep learning techniques networks. The training set is used to train all models, and the testing data set is used to evaluate performance for various setting for hyper- parameters.

6.7.1 Performance of Random Forest with PCA

Confusion Matrix

Figure 6.10 shows the confusion matrix of combined random forest and PCA model.

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.8,and the performance parameters for evaluating the model is represented in table 6.9.

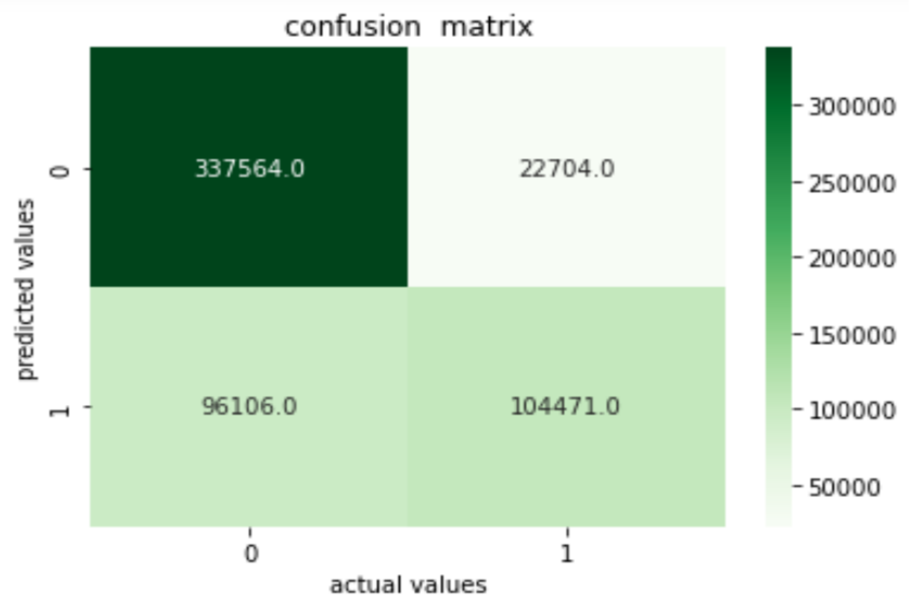

Figure 6.10: Confusion matrix of Random Forest with PCA

Table 6.8: Values from confusion matrix of Random Forest-PCA model

TP	TN	FP	FN
337564	104471	22704	96106

Table 6.9: Performance parameters of Random Forest-PCA model

Accuracy (%)	Precision	Recall	F1 score
78.81	0.936	0.77	0.849

6.7.2 Performance of Random Forest without PCA

Confusion Matrix

Figure 6.11 shows the confusion matrix of random forest model.

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.10,and the performance parameters for evaluating the model is represented in table 6.11.

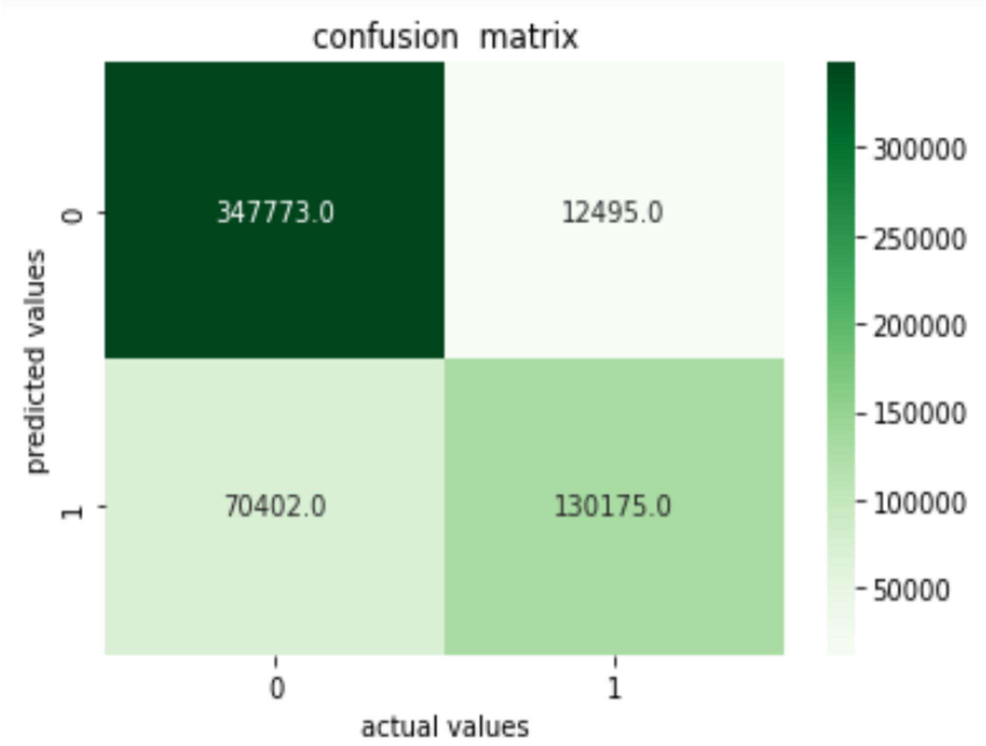

Figure 6.11: Confusion matrix of Random Forest with out PCA

Table 6.10: Values from confusion matrix of Random Forest model

TP	TN	FP	FN
347773	130175	12495	70402

Table 6.11: Performance parameters of Random Forest model

Accuracy (%)	Precision	Recall	F1 score
85.219	0.965	0.831	0.8934

6.7.3 Performance of Vanilla LSTM

Model Accuracy Graph

Accuracy-epoch plot of vanilla LSTM is shown in figure 6.12.The model obtained an accuracy of only 76.7% at the end of iteration.

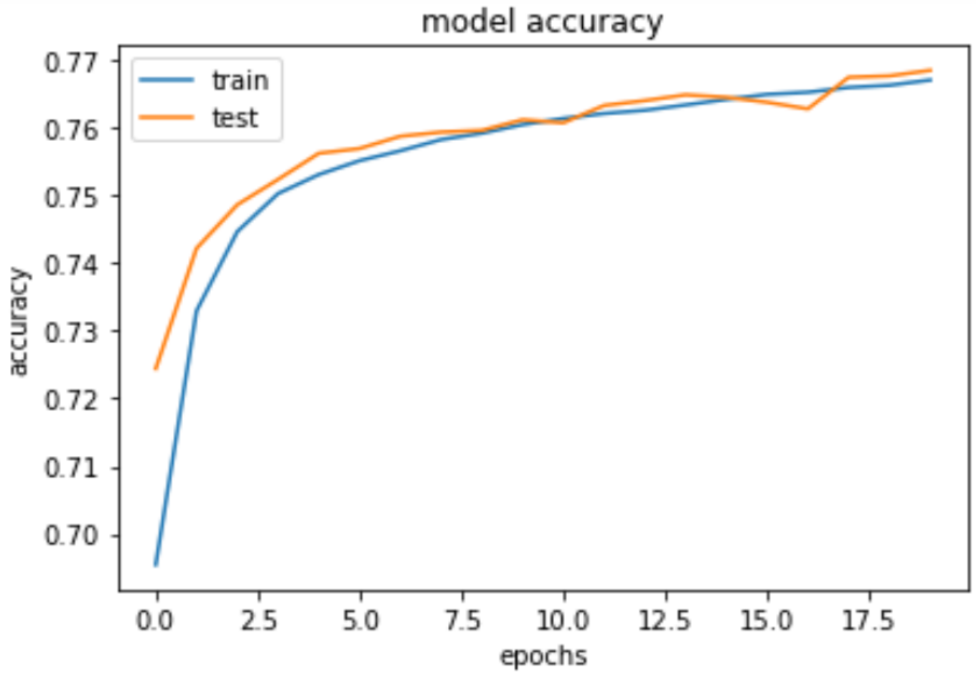

Figure 6.12: Accuracy Graph of Vanilla LSTM model

Model Loss Graph

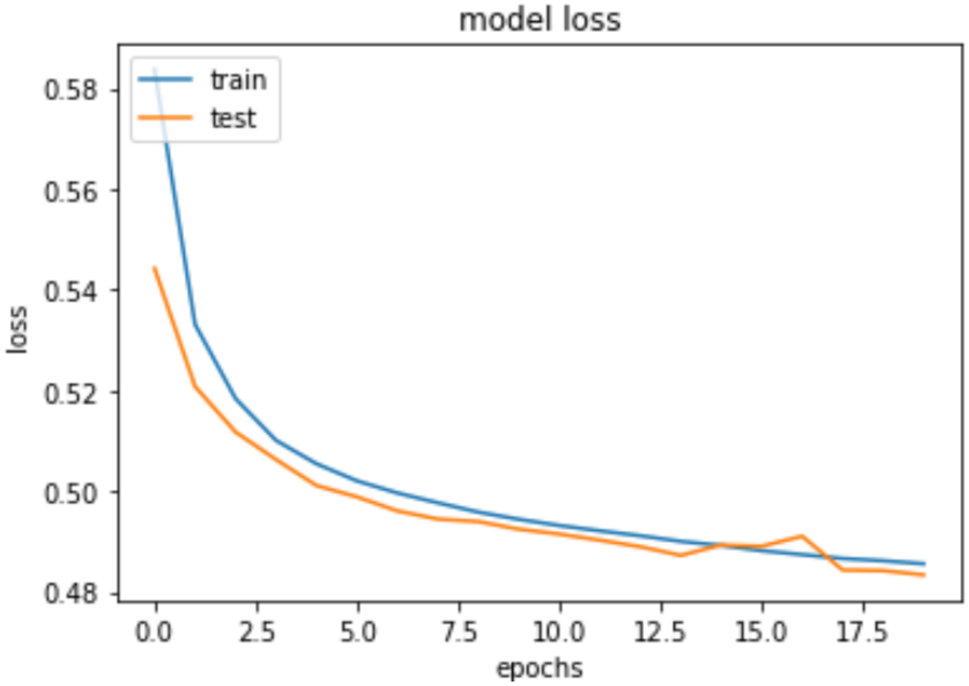

Figure 6.13: Loss Graph of Vanilla LSTM model

Figure 6.13 shows the loss-epoch plot of Vanilla LSTM model.The loss function started from 0.5 and at the end of epochs it decreased to only 0.48.

Confusion Matrix

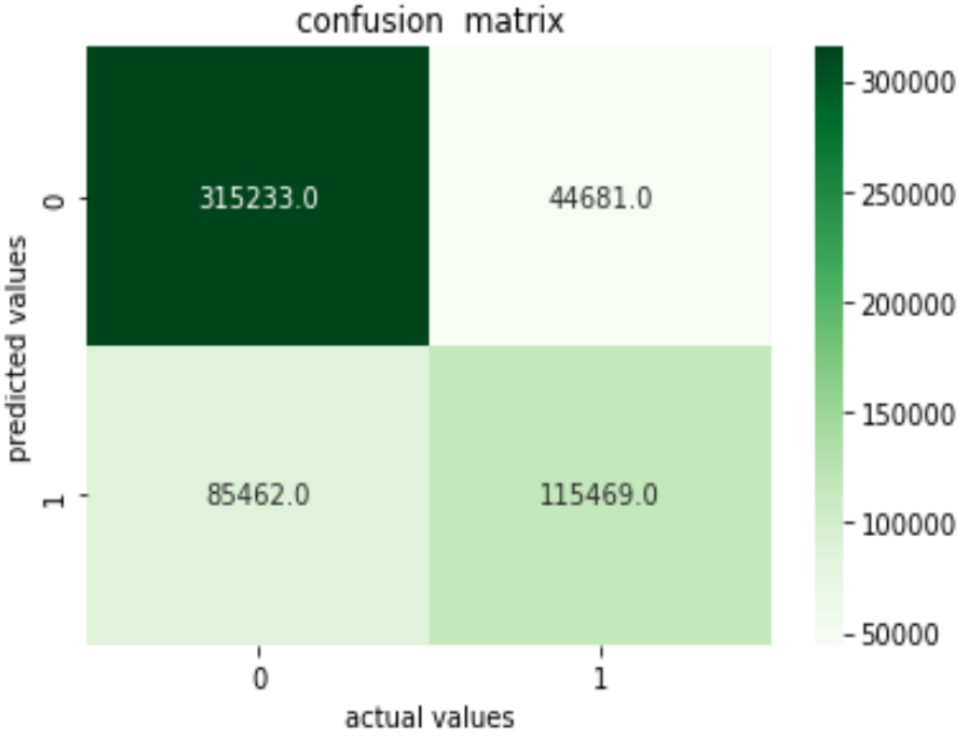

Figure 6.14: Confusion matrix of Vanilla LSTM model

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.12,and the performance parameters for evaluating the model is represented in table 6.13.

Table 6.12: Values from confusion matrix of Vanilla LSTM model

TP	TN	FP	FN
315233	115469	44681	85462

Table 6.13: Performance parameters of Vanilla LSTM model

Accuracy (%)	Precision	Recall	F1 score
76.795	0.875	0.786	0.828

6.7.4 Performance of GRU

Model Accuracy Graph

Accuracy-epoch plot of GRU is shown in figure 6.15.The model obtained an accuracy of 91% at the end of iteration.

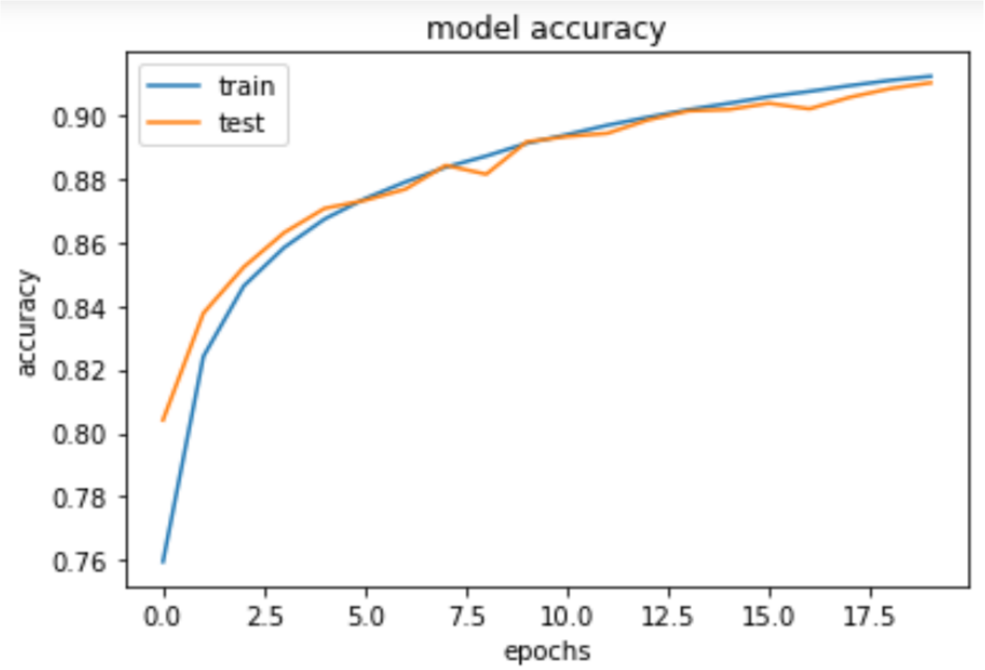

Figure 6.15: Accuracy Graph of GRU model

Model Loss Graph

Figure 6.16 shows the loss-epoch plot of Vanilla LSTM model.The loss function started from 0.5 and at the end of epochs it decreases to 0.2.

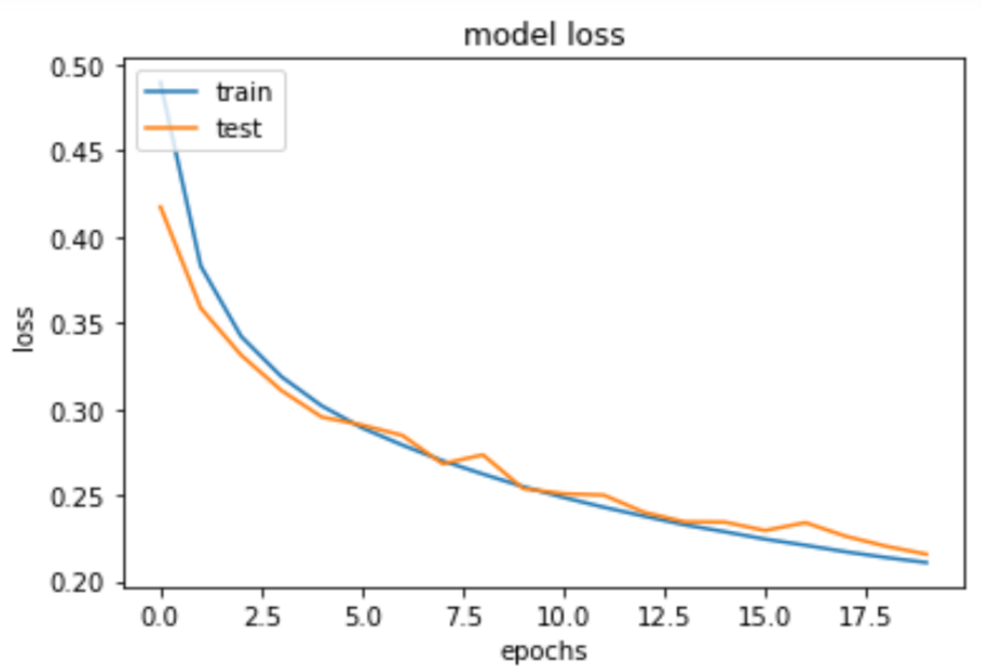

Figure 6.16: Loss Graph of GRU model

Confusion Matrix

Confusion matrix of GRU model is shown in figure 6.17

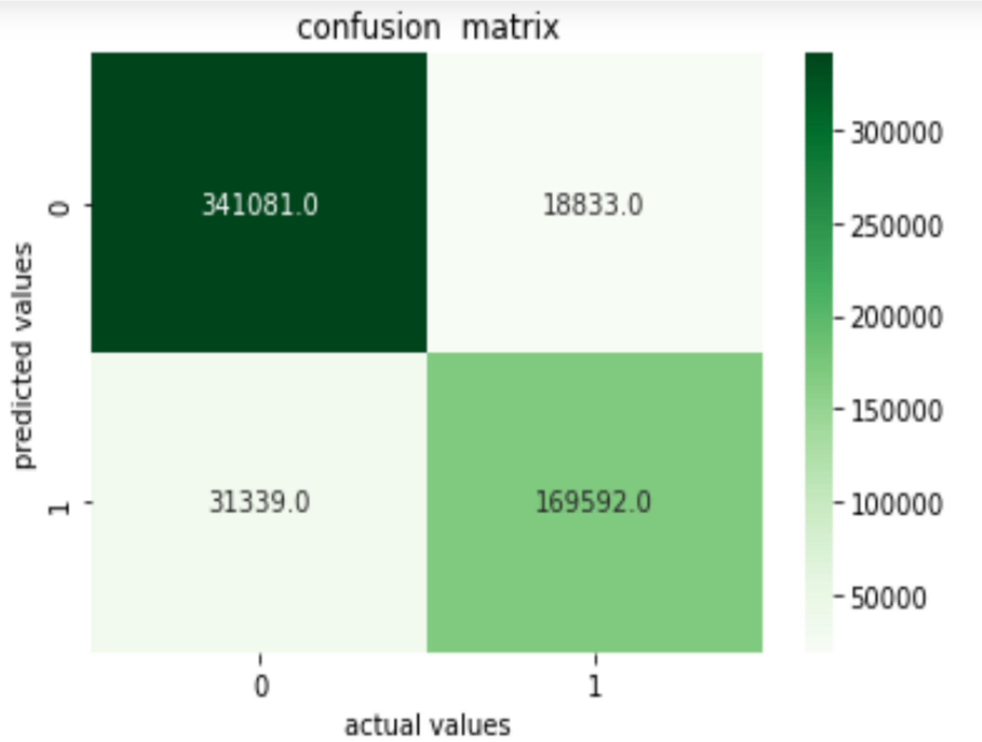

Figure 6.17: Confusion matrix of GRU model

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.14,and the performance parameters for evaluating the model is represented in table 6.15.

Table 6.14: Values from confusion matrix of GRU model

TP	TN	FP	FN
341081	169592	18833	31339

Table 6.15: Performance parameters of GRU model

Accuracy (%)	Precision	Recall	F1 score
91.05	0.947	0.915	0.931

6.7.5 Performance of BiLSTM

Model Accuracy Graph

Accuracy-epoch plot of BiLSTM is shown in figure 6.18.The model obtained an accuracy of 95.87% at the end of iteration.

Model Loss Graph

Figure 6.19 shows the loss-epoch plot of BiLSTM model.The loss function started from 0.45 and at the end of epochs it decreases to 0.1.

Confusion Matrix

Confusion matrix of BiLSTM model is shown in figure 6.20

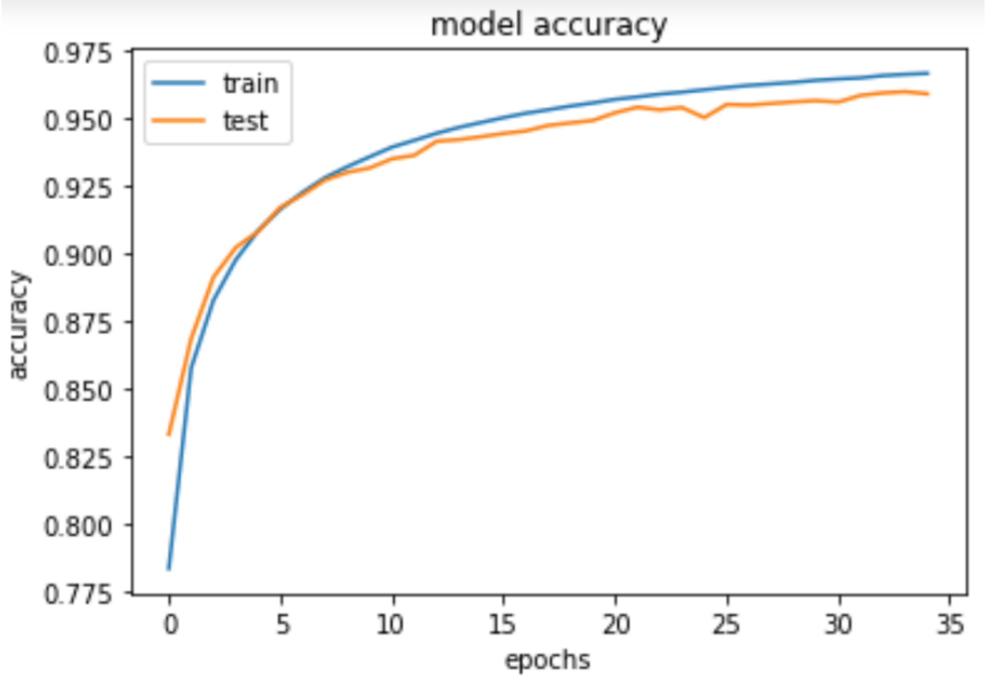

Figure 6.18: Accuracy Graph of BiLSTM model

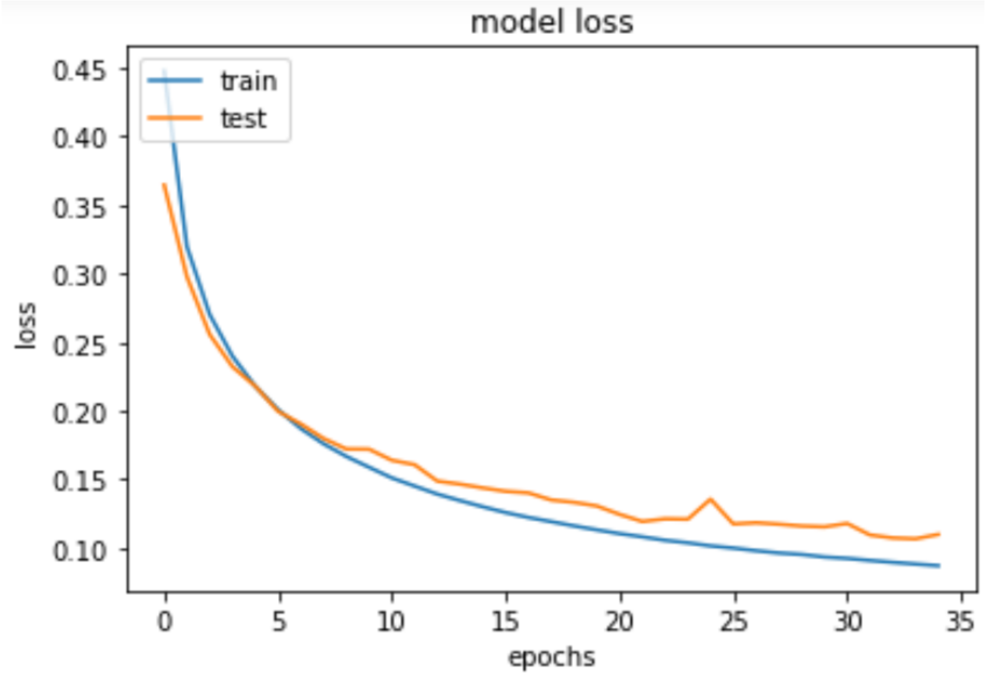

Figure 6.19: Loss Graph of BiLSTM model

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.16,and the performance parameters for evaluating the model is represented in table 6.17.

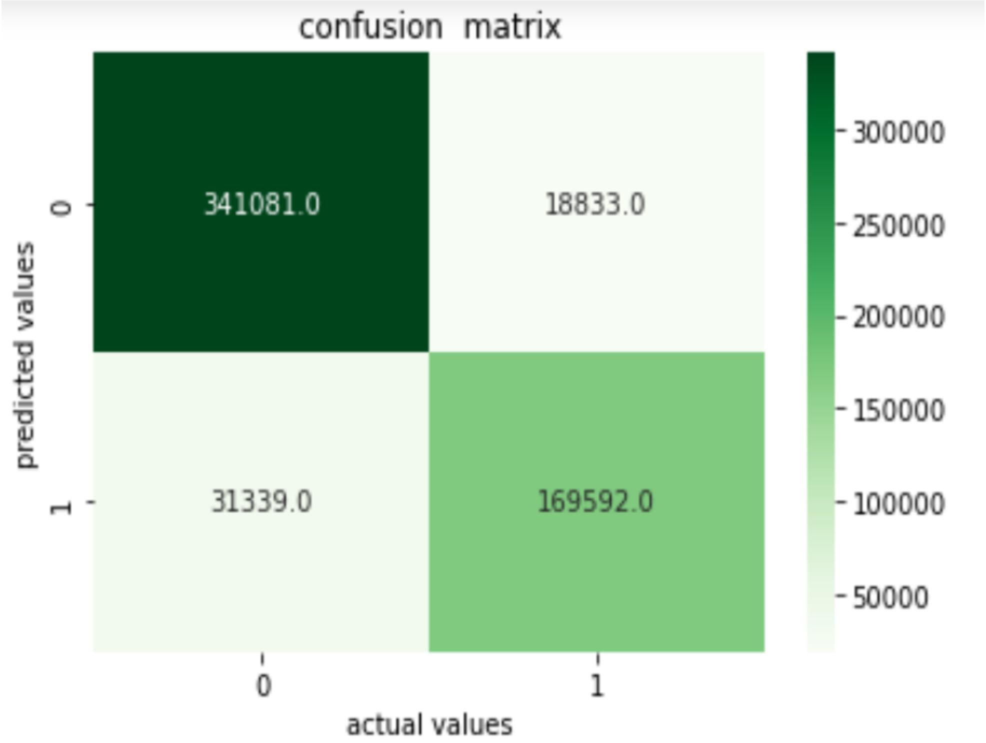

Figure 6.20: Confusion matrix of BiLSTM model

Table 6.16: Values from confusion matrix of BiLSTM model

TP	TN	FP	FN
349582	188144	10332	12787

Table 6.17: Performance parameters of BiLSTM model

Accuracy (%)	Precision	Recall	F1 score
95.87	0.9712	0.964	0.967

6.8 Performance of Ensembled Models

Confusion Matrix

Confusion matrix of Ensembled model is shown in figure 6.21.

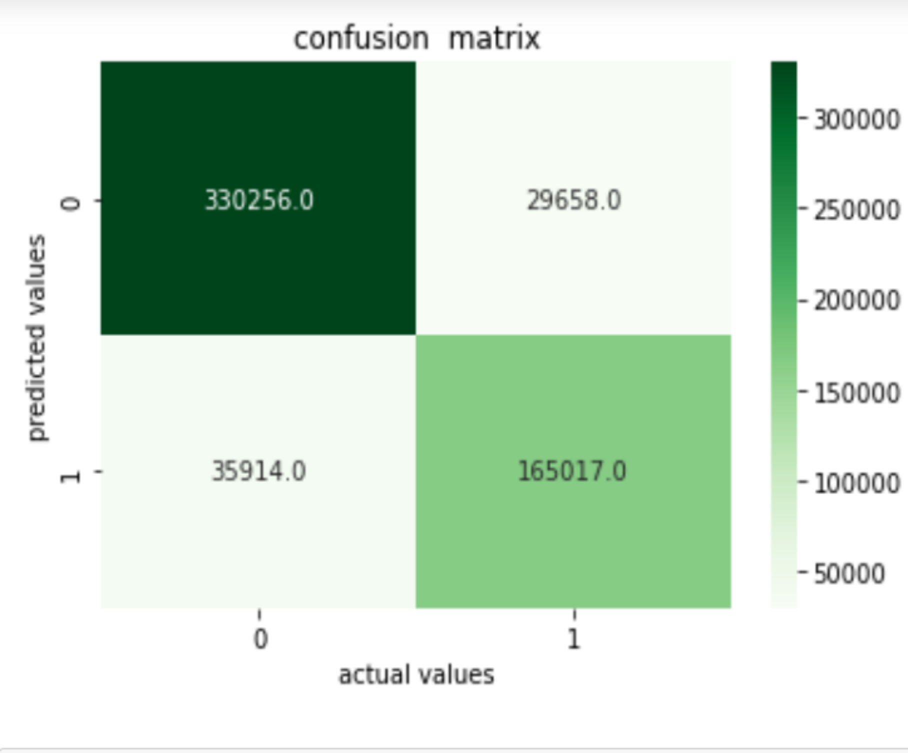

Figure 6.21: Confusion matrix of Ensembled model

Results from Confusion Matrix

TP,TN,FP,FN obtained from confusion matrix is shown in table 6.18,and the performance parameters for evaluating the model is represented in table 6.19.

Table 6.18: Values from confusion matrix of Ensembled model

TP	TN	FP	FN
330256	165017	29658	35914

Table 6.19: Performance parameters of Ensembled model

Accuracy (%)	Precision	Recall	F1 score
88.30	0.91	0.90	0.91

6.8.1 Design Parameters

Table 6.20: PROPOSED MODEL ARCHITECTURE

Proposed model	Model	Activation function	Criterion	Topology	Epoch	Dropout
Algorithm 1	PCA+Random Forest		Gini,Entropy			0
	Random Forest		Gini,Entropy			0
Algorithm 2	GRU	Relu,sigmoid		64-100-1	10-50	0.0-0.4
Algorithm 3	Vanilla LSTM	Relu,sigmoid		64-100-1	10-50	0.0-0.4
Algorithm 4	BiLSTM	Relu,sigmoid		64-100-1	10-50	0.0-0.4

Table 6.21: DEEP LEARNING MODEL DESIGN-HYPER PARAMETERS

Proposed algorithm	Model	Optimizer	Batch size	Loss function
Algorithm 2	GRU	Adam,SGD	50-200	Mean Squared Error, Binary Cross Entropy
Algorithm 3	Vanilla LSTM	Adam,SGD	50-200	Mean Squared Error, Binary Cross Entropy
Algorithm 4	BiLSTM	Adam,SGD	50-200	Mean Squared Error, Binary Cross Entropy

6.8.2 performance of Proposed Models

Table 6.22: BEST RESULTS FROM PROPOSED MODELS

Proposed algorithm	Model	Accuracy (%)	Execution time	Precision (%)	Recall (%)	F1 score (%)
Algorithm 1	PCA+Random Forest	78.81	30:00	93.6	77.8	84.97
	Random Forest	85.219	45:00	96.53	83.164	89.34
Algorithm 2	GRU	91.05	06:00:00	94.76	91.58	93.14
Algorithm 3	Vanilla LSTM	76.795	06:00:00	87.58	78.671	82.88
Algorithm 4	BiLSTM	95.87	10:00:00	97.129	96.471	96.798
Algorithm 5	Ensembled model	88.30	2:00:00	91.75	90.68	91.14

Analysing table 6.7 and table 6.22 ,it's clear that BiLSTM performs as the best classifier when compared with other models.It shows that, in comparison to the current approaches for classification , the proposed BiLSTM-based algorithm produces the higher classification results (95.87%).

Graphical performance comparison of all implemented and proposed models

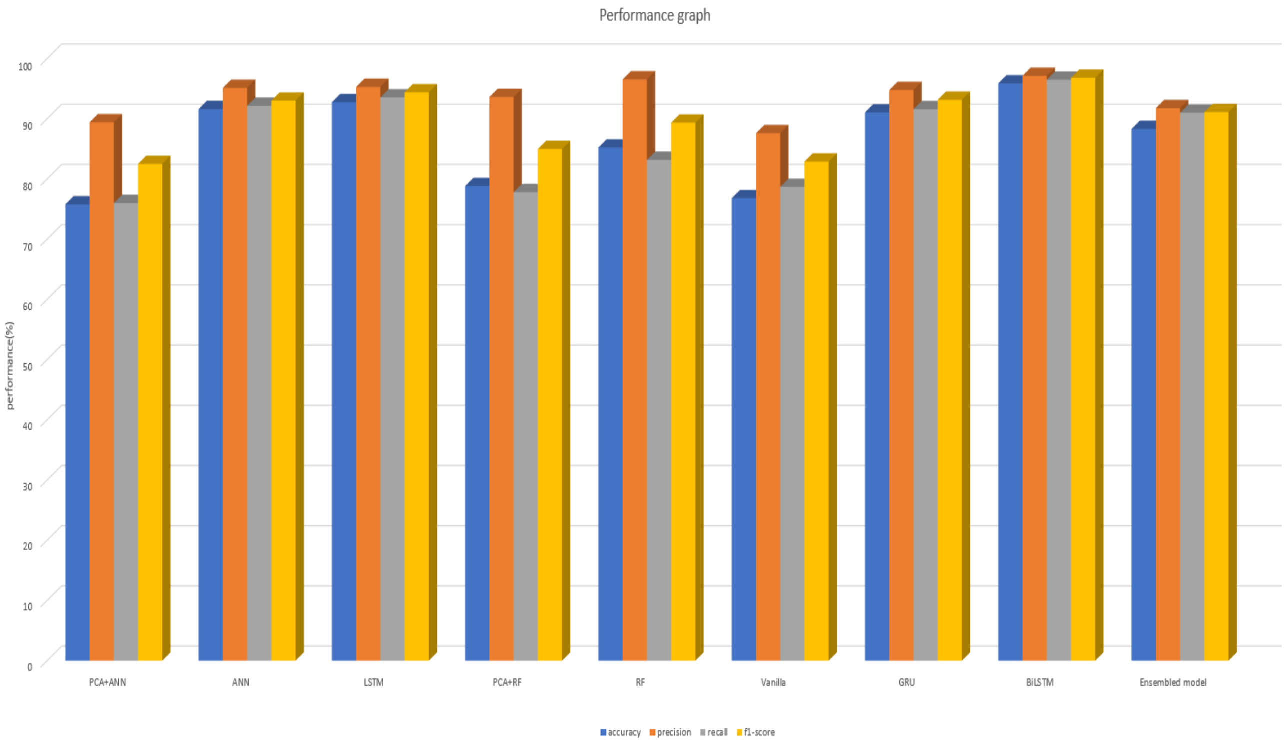

Figure 6.22: comparison chart

6.9 Conclusion

This chapter verified the result obtained from the different experimentation conducted.Different parameters are used to analyse the model's performance. Moreover, to emphasise the superiority of the suggested work, performance of all models are tabulated.

Chapter 7

CONCLUSION AND FUTURE SCOPE

A deep learning algorithm is proposed in this research that would improve the accuracy of alcoholic and non-alcoholic EEG signals classification. This study sought to provide an effective method for effectively categorising EEG data into two classes, alcoholic and non-alcoholic. One machine learning algorithm , three deep learning based methodologies and one ensemble model were suggested in this study: Algorithm 1 based on PCA and Random Forest , while Algorithm 2, Algorithm 3 and Algorithm 4 is based on deep learning models like GRU,Vanilla LSTM and BiLSTM networks respectively and algorithm 5 is ensemble model of random forest and XG boost. On the UCI Alcoholic EEG dataset, the suggested algorithm's performances were evaluated in terms of accuracy,recall,precision and f1-score. Our study demonstrated that the BiLSTM approach (Algorithm 4) has significantly improved performance (95.87% accuracy,96.79% f1-score and 96.47%recall), indicating that BiLSTM is a strong and stable classifier for EEG data,which can be used in medical applications. The outcome of other algorithm's is obviously lower than the output of the BiLSTM model. However, the BiLSTM was not fast and its run time was about 10 hours,but it provides a better accuracy when compared with other deep learning models of previous study.The future scope of the project includes investigation of better models for classification.

REFERENCES

[1] Acharya, U. R., S, V., Bhat, S., Adeli, H., Adeli, A. (2014). "Computer-aided diagnosis of alcoholism-related EEG signals". *Epilepsy Behavior*, 41, 257–263.

[2] Stickel, F., Schuppan, D., Hahn, E. G., Seitz, H. K. (2002). "Cocarcinogenic effects of alcohol in hepatocarcinogenesis". Gut, 51(1), 132–139. doi:10.1136/gut.51.1.132

[3] World Health Organization (WHO). *Alcohol Fact Sheet, Geneva (WHO, 2011).* Accessed: Sep. 21, 2018.

[4] *National Institute of Health, National Institute on Alcohol Abuse and Alcoholism*, U.S. Dept. Health Hum. Services, Washington, DC, USA, 2004.

[5] S. F. Tapert, G. G. Brown, S. S. Kindermann, E. H. Cheung, L. R. Frank, and S. A. Brown, "FMRI measurement of brain dysfunction in alcohol dependent young women," *Alcoholism, Clin. Exp. Res.*, vol. 25, no. 2, pp. 236–245, Feb. 2001.

[6]] S. Siuly, V. Bajaj, A. Sengur, and Y. Zhang, "An advanced analysis system for identifying alcoholic brain state through EEG signals," *Int. J. Autom. Comput.*, vol. 16, no. 6, pp. 737–747, Dec. 2019.

[7] Varun Bajaj, Yanhui Guo, Abdulkadir Sengur, Siuly Siuly, and Omer F Alcin. "A hybrid method based on time–frequency images for classification of alcohol and control eeg signals". *Neural Computing and Applications*, 28(12):3717–3723, 2017

[8] Oliver Faust, Wenwei Yu, and Nahrizul Adib Kadri. Computer-based identification of normal and alcoholic eeg signals using wavelet packets and energy measures. *Journal of Mechanics in Medicine and Biology*, 13(03):1350033, 2013.

[9] U Rajendra Acharya, S Vinitha Sree, Subhagata Chattopadhyay, and Jasjit S Suri." Automated diagnosis of normal and alcoholic eeg signals. *International journal of neural systems*, 22(03):1250011, 2012

[10]] Oliver Faust, RU Acharya, Alastair R Allen, and CM Lin. Analysis of eeg signals during epileptic and alcoholic states using ar modeling techniques. *Irbm*, 29(1):44–52, 2008.

[11] Arti Anuragi and Dilip Singh Sisodia. "Alcohol use disorder detection using eeg signal features and flexible analytical wavelet transform". *Biomedical Signal Processing and Control*, 52:384–393, 2019.

[12] Cindy L Ehlers, James Havstad, Dean Prichard, and James Theiler. Low doses of ethanol reduce evidence for nonlinear structure in brain activity.*Journal of Neuroscience*,1998.

[13] Mohammad Reza Nazari Kousarrizi, Abdolreza Asadi Ghanbari, Ali Gharaviri, Mohammad Teshnehlab, and Mahdi Aliyari. Classification of alcoholics and non-alcoholics via eeg using svm and neural networks. In 2009 *3rd International Conference on Bioinformatics and Biomedical Engineering*, pages 1–4. IEEE, 2009.

[14] Hua Wang, Yanchun Zhang, et al. An efficient framework for the analysis of big brain signals data. In *Australasian database conference*, pages 199–207. Springer, 2018

[15] Shrey Agarwal and Muhammad Zubair. Classification of alcoholic and non-alcoholic eeg signals based on sliding-ssa and independent component analysis. *IEEE Sensors Journal*, 21(23):26198–26206, 2021.

[16] Leila Farsi, Siuly Siuly, Enamul Kabir, and Hua Wang. "Classification of alcoholic eeg signals using a deep learning method". *IEEE Sensors Journal*, 21(3):3552–3560, 2020.

[17] C. Blake and C. Merz, “UCI repository of machine learning databases,” Dept. Inf. Comput. Sci., Univ. California, Irvine, CA, USA, Tech. Rep., 1998.

[18] L. Smith, “A tutorial on principal components analysis,” Dept. Comput. Sci., Univ. Otago, Dunedin, New Zealand, Tech. Rep. OUCS-2002-12, 2002.

[19] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997

[20] S. Siuly, Y. Li, and Y. Zhang, *EEG Signal Analysis and Classification: Techniques and Applications* (Health Information Science). Cham, Switzerland: Springer, 2016

[21] Alotaiby, T., El-Samie, F.E.A., Alshebeili, S.A. et al. A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process. 2015, 66 (2015).*

[22] Dietterich, T.G. (2000). "Ensemble Methods in Machine Learning. In: Multiple Classifier Systems". MCS 2000. *Lecture Notes in Computer Science, vol 1857. Springer*, Berlin, Heidelberg.