An Ensemble based Framework for Eye State Prediction from EEG data

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
In
Computer Science and Engineering

Submitted By
Sukhdeep Kaur
(Roll No. 801632050)

Under the supervision of:
Dr. Shalini Batra
Associate Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

June 2018

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "***An Ensemble based Framework for Eye State Prediction from EEG data***", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of ***Dr. Shalini Batra*** and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Sukhdeep Kaur.

Sukhdeep Kaur

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Dr. Shalini Batra

Associate Professor,

Computer Science and Engineering Department

Countersigned by

(Dr. Maninder Singh)

Head

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology

Patiala

(Dr. S S Bhatia)

Dean (Academic Affairs)

Thapar Institute of Engineering and Technology

Patiala

ACKNOWLEDGEMENT

I would first like to express my sincere gratitude to my thesis supervisor **Dr. Shalini Batra,** Associate Professor, Computer Science and Engineering Department, Thapar Institute of Engineering and Technology for the continuous support of my M.E. reasearch work, for her motivation, patience, enthusiasm, and immense knowledge. She has been a constant source of inspiration for me. I could not have imagined having a better supervisor and mentor for my research work. Her guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank **Dr. Maninder Singh**, Head CSED, and **Dr. Ashutosh Mishra,** PG Coordinator, for their constant motivation and encouragement. I am also thankful to non-teaching staff of the institute for their support and help.

Finally, I would like to express my sincere and deep gratitude to my parents, family members and friends for their love, care, encouragement, and support.

Sukhdeep Kaur

ABSTRACT

Since eyes are an important organ of human beings, eye gestures can be widely used in various smart systems, especially in identifying brain activity pattern. Electric signals are transmitted between brain cells for transfer of data, therefore using Electroencephalography (EEG), electrical activity of the brain can be recorded. With the help of small electrodes placed on human scalp, these electric signals can be captured. Based on this EEG, state of eye i.e. open or closed is predicted. Eye State Prediction has application in many fields such as driver drowsiness prediction using sequence of eye states, stress detection, military scenarios, etc. It plays a vital role in medical field also where it can be used as medium of communication for patients who are paralyzed or severely handicapped. It can also be used for controlling computer systems using eye gaze. In the previous work done in this field, data mining techniques and classifiers have been applied to the EEG dataset, however, hybridization (or ensembling) in addition to individual classifiers yields better results. In this study, we have developed an ensemble based model based on majority voting technique and compared its performance with different classifiers on the basis of various performance parameters such as accuracy, sensitivity, and specificity, and shown that ensemble based model performs better than other classifiers.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
KDD	Knowledge Discovery in Databases
UCI	University of California
CART	Classification and Regression Trees
kNN	k-Nearest Neighbors
SVM	Support Vector Machine
GLM	Generalized Linear Model
IG	Information Gain
GBM	Gradient Boosting Machines
TD	Temporal Difference
TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives

Chapter 1

Introduction

Many scientists consider human brain as a black box. Although majority of the brain's workings is still a mystery, some phenomenon can be modelled and explained. With the detection of electrochemical signals and blood flow, the brain's activity can be measured. Electroencephalography (EEG) is an electrophysiological checking strategy to record electrical action of the cerebrum, which is produced by firing of neurons in the brain. Brain Machine Interface (BMI) and Brain Computer Interface (BCI) are EEG based communication systems that use the brain signal and perform many tasks by means of brain electrical activity. We present an ensemble based framework for eye state detection using EEG signals and include processes like feature extraction, attribute selection and classification. The requirement for precise and quick arrangement and forecast of brain activity is winding up more critical with the approach of mind PC interfacing technology because from touch activated and voice controlled devices to the next generation of user interface, control system is activated through thought process. Mental activity leads to changes of electrophysiological signals like the EEG. Since eyes are an important organ of human beings, their gestures can be used for controlling various smart systems and identifying brain activity pattern. Eye state has been utilized in various fields of application such as operating computer by an extremely physically challenged individual [13]. Gestures involving eye state can also support various applications in the digital world. It also has application in computer games and military scenarios. Eye state movements can be utilized to control a cursor by moving one's look from one edge of the computer display to the other [20], or to control virtual keyboards through patterned blinks [6]. This data can also be utilized in medical applications such as to distinguish between patients with attention deficit hyperactivity order (ADHD) and biploar mood disorder (BMD) [21].

Eye State prediction can be used in medical field, where eye movements can be only means of communication for patients with pseudocoma, as such patients encounter complete loss of motion of all voluntary muscles aside from those controlling eyes. EEG data can be also be used for determining, monitoring and predicting driver fatigue, as eye states can reflect level of driver's fatigue. Since driver drowsiness has become main cause of traffic accidents, successfully predicting eye state can reduce rate of traffic accidents by monitoring driver's drowsiness.

Eye State Prediction can also be used for tracking emotions, neurological disorder detection, and sleep analysis. It has also been effectively applied in the field of sleep-walking state prediction, epileptic seizure detection, and detecting stress with help of eye state frequency.

1.1 Data Mining

Huge volume of data is present in the Information Industry which is of no use until changed into valuable information. Here data mining comes into picture. It is a way towards extracting useful information from huge datasets and then discovering patterns in those datasets through machine learning and statistics. It turns raw data into understandable structure for further use. Apart from this, it also involves data pre-processing and data management.

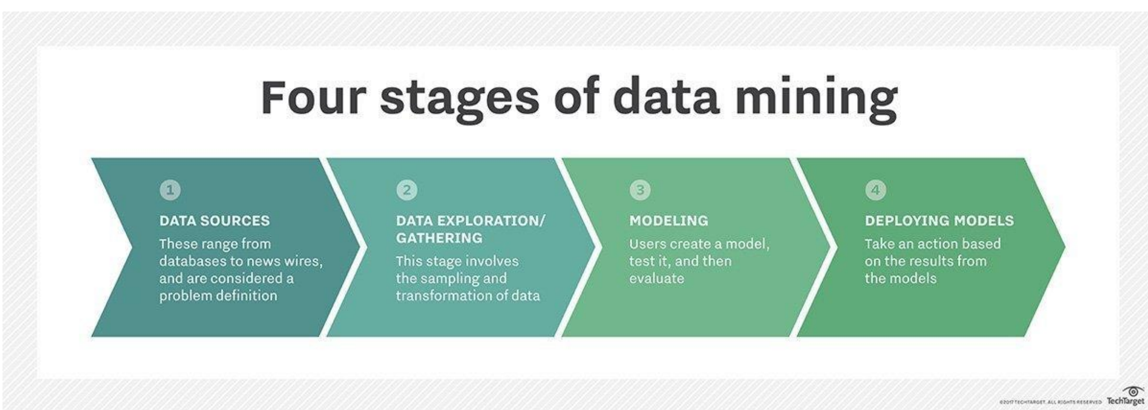

Fig 1.1 : Data mining stages [26]

Key features of data mining include:

- Creation of decision-oriented information
- Focus on databases and large data sets for analysis
- Automatic discovery of patterns
- Prediction of likely outcomes
- Clustering based on finding

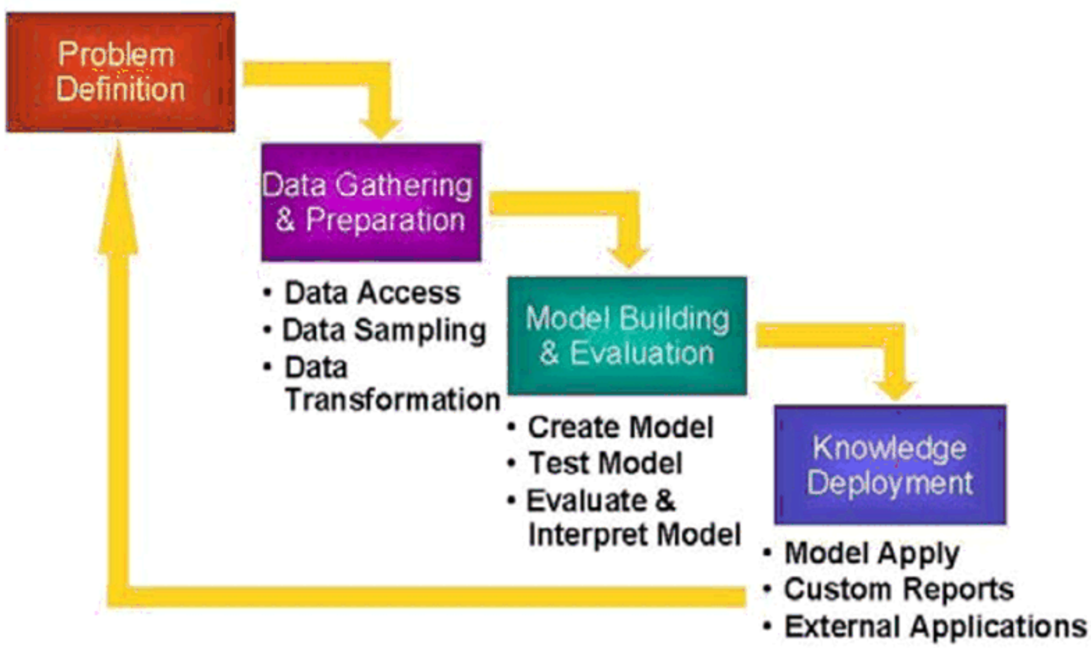

Fig 1.2 : Data Mining Process [27]

Data Mining is used in applications such as Customer Retention, Production Control, Fraud Detection, Market Analysis, Corporate Analysis, Risk Management, and Science Exploration.

Following data mining tasks are associated in eye state prediction using EEG data:

- **Descriptive Tasks:** In this task, main objective is to derive patterns using correlations, clusters, trends, etc. It manages general properties of information in database. Some of the descriptive functions include:

- Mining of clusters
- Mining of correlations
- Mining of associations
- Concept/Class description
- Mining of frequent patterns

- **Predictive Tasks:** Objective of this task is to predict value of a specific attribute also known as dependent or target variable based on the values of other attributes also known as independent or explanatory variables. It includes 2 different types of tasks, i.e. classification and regression.
- **Anomaly Detection :** It is the identification of unusual data records or those records whose characteristics are different from rest of the data in the dataset. Its objective is to avoid incorrect labelling.
- **Cluster Analysis :** The objective of this task is to find structures and groups in the data that are similar in one way or the another.

1.2 KDD (Knowledge Discovery in Databases)

It is the process of finding useful information from huge volume of data.

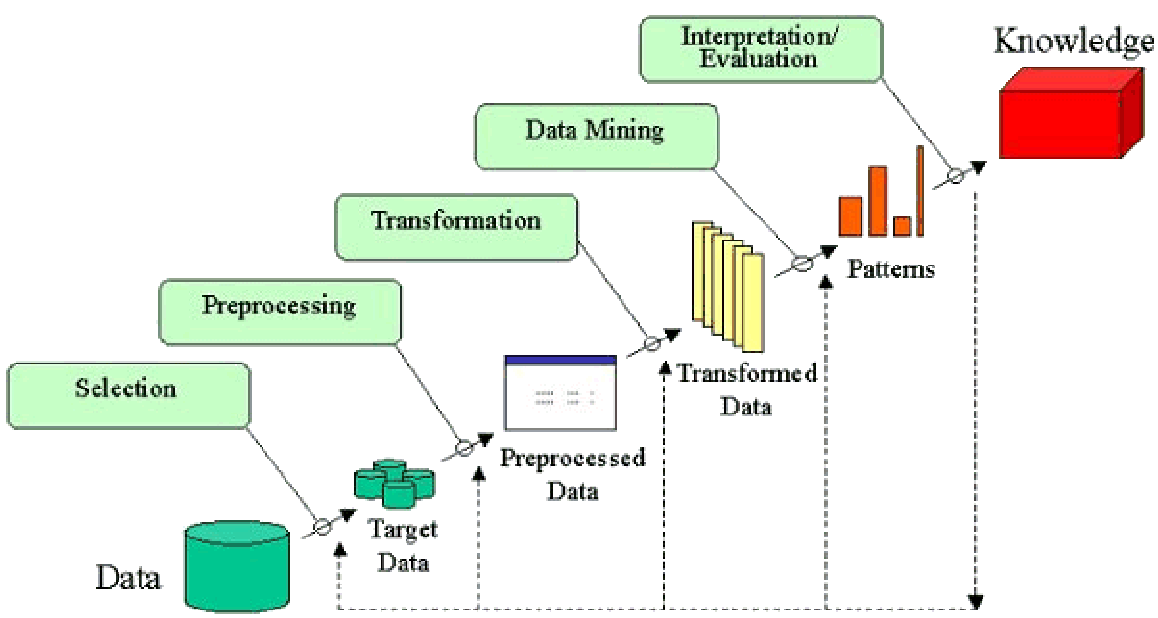

Fig 1.3 : Steps of KDD [28]

Steps of KDD are as:

- Understanding of
    - The domain of application
    - End-user goals
    - Prior knowledge relevant to application
- Creation of target data set
    - Selection of dataset
    - Focus on subset of data samples, or variables to discover knowledge
- Data cleaning and pre-processing
    - Removing outliers or noise
    - Collection of sufficient information to account for noise
    - Handle missing data fields
    - Time sequence information accounting
- Data reduction and projection
    - According to goal of task, represent data by finding important and useful features
    - Reducing number of variables by using transformation methods and dimensionality reduction
- Choosing the data mining task
    - Whether it is regression, classification, clustering, etc.
- Choosing data mining algorithm
    - Selection of technique to discover patterns in the data
    - Decision of appropriate parameters
- Data mining
- Interpretation of discovered patterns
- Consolidation of discovered knowledge

KDD also includes pre-processing, projections, encoding schemes, and sampling of data prior to data mining step.

1.3 Thesis Outline

This thesis is categorized into 6 chapters:

- **Chapter 1 :** This chapter gives brief introduction to machine learning, data mining concepts and their application in various fields.
- **Chapter 2 :** This provides literature survey of the research area and brief explanation of previous works done in that area.
- **Chapter 3 :** This chapter gives problem statement in brief along with objectives of research.
- **Chapter 4 :** This chapter describes proposed approach to fulfil the objectives of research.
- **Chapter 5 :** Results generated from the proposed approach are presented in this chapter, and it also discusses the performance of various models applied to the data set.
- **Chapter 6 :** This chapter gives conclusion and future scope of the research work.

Chapter 2

Literature Review

This chapter provides summary of all the work that has been done in this field of eye state prediction.

Today, Data Mining approach is widely used in many applications. In large data sets, data mining approach is used to discover patterns and then making predictions on that data set at the intersection of machine learning, database systems, and statistics. In this, intelligent techniques are applied to extract data patterns. Its primary objective is to obtain information from the data set and convert it into comprehensible structure for further use. KDD (Knowledge Discovery in Databases) is a procedure of finding knowledge in data, which focuses on the "high-level" application of specific data mining methods, and has become a popular research tool. Its unifying goal is to extract knowledge from large databases. As data is increasing at an exponential rate, need for KDD and data mining is also increasing.

In the recent years, machine learning algorithms and data mining techniques have been widely used in various areas of science and engineering, and has played an important role in tackling the data overload in various fields.

Rajesh *et al*.[2] used C4.5 Decision Tree Algorithm for eye state prediction. Out of total 14980 instances, they utilized 4000 instances, in which 2000 instances were there for eye state being open and close each. In dataset, there were 14 attributes, which they ranked according to their significance, and they used only 10 attributes having more significance. With 13 attributes, they yielded maximum accuracy of 91.6% but with 10 attributes, accuracy came was 90.72%. Since there is not much difference in accuracy, it is better to use less attributes. They verified classification by ten-fold cross validation method. This paper presented eye state prediction with the use of C4.5 Decision Tree Algorithm with 83.5% accuracy which is acceptable for various practical applications.

Gopika *et al.*[14] proposed an approach which used statistical features along with k-Nearest Neighbor (kNN) and Support Vector Machine (SVM). After feature extraction, they selected 13 features and gave as input to kNN (k=7) and SVM (with RBF kernel) classifiers. It yielded mean accuracy of 77.92%. They used 10-fold cross validation for classification.

Neha *et al.*[22] applied four different machine learning algorithms i.e., SVM Linear, SVM Polynomial, SVM RBF, and K star using Weka 3.6.4 machine learning software. They adopted for 10-fold cross validation for random sampling of the training and testing data set. By analyzing the performance of all models on the basis of accuracy, F-measure, recall, and precision, they gave result that K star classifier yielded maximum accuracy of 97.30%.

Oliver *et al.*[4] compared 42 different machine learning algorithms. They divided attributes into 2 groups. One group having minimum decreases when eyes open, and other group having maximum increases in the same event. On the basis of comparison, they concluded that K Star is best performing classifier with 97.3% accuracy and 2.7% error rate.

Cameron *et al.*[3] developed three ensemble learners: a rotational forest in which random forests are implemented as base classifiers (RBF), a rotational forest in which J48 trees are implemented as base classifiers, and then boosted with adaptive boosting, and ensemble of RRF model. After comparison of above, they concluded that ensemble structure of K start classifiers and RBF yielded highest accuracy of 97.4%, whereas RBF model with 10 forests of 300 trees each generated accuracy of 95.1%, where each tree's decision node was based on three features. Accuracy with adaptive boosted RJ48F was 97.2% using 10 J48 trees for 50 iterations.

Reddy and Behera[16] used deep architectures for eye state prediction. They designed deep belief networks based unsupervised learning, drop-out masks on deep neural networks, and a multi-layered neural network with ReLU and drop-out. They used 10-fold Cross Validation, Leave-one out Cross validation (LOOCV) for partitioning of data. Their method achieved accuracy of 97.5% and was faster than previous methods as it achieved classification speed of at-least 1000 samples per second. They concluded that neural classifiers are faster than instance based classifiers such as K star etc., and deep learning reduces time for convergence and improves accuracy significantly.

Norma *et al.*[23] used Circular Hough Transform and Haar Cascade Classifer for eye state prediction. It yielded an accuracy of 96.96%. Fuangkaew and Patanukhom[24] used SVM classifiers for eye state recognition with accuracy of 95.37%. Cui *et al.*[25] yielded 98% accuracy of eye state prediction by using an Adaboost based cascaded classifier.

Chapter 3

Problem Statement

Data mining has a vital role in prediction problems. The premise of machine learning is revelation of new information in terms of patterns or rules from huge amount of data. Eye State Prediction has application in many fields such as controlling cursor with gaze, operating computers by physically challenged persons, to predict dizziness during driving, detecting stress using eye blinks and brain activity, etc. With data mining approach, we get to know which attributes and features are more important for the prediction purpose. The research work presented in this thesis takes the challenge to improve prediction model. More accurately we predict state of eye, more accurately its other applications will be predicted. Summarizing important research functions as:

- In what way can we use data mining techniques in prediction problems and to identify their performance?
- In what way classification techniques help in prediction problems?

3.1 Objectives of Research

Data mining plays an important role in finding the existing relationships between different variables and attributes. The main objective of this research work is the development of prediction model using various classifiers, and then evaluating its performance. Following objectives are laid out for this work:

- Study of data mining, various classification techniques in machine learning, and advantage of combining the results of more than one classifier.
- To propose an ensemble framework for eye state prediction.
- To evaluate the performance of proposed ensemble framework with other models on different parameters such as sensitivity, specificity, and accuracy

Chapter 4

Proposed Approach: Ensemble based Approach

4.1 Machine Learning

Machine Learning is an AI (Artificial Intelligence) technique that is about setting systems to the task of searching through data to look for patterns and altering actions accordingly, with an ultimate aim to enable software applications to become more precise without being specifically programmed. Its fundamental theory is to build programs that can get input data and utilize statistics to foresee an output while updating outputs as new data becomes available, hence improving automatically through experience. It is the science of making our computers to act like humans, and has been central to AI research since the field's inception. It is based on algorithms that can find out the way to perform significant tasks by concluding from examples, and that can gain from data without depending on rule-based programming. Its primary target is to generalize beyond the training samples, i.e. to effectively expound data it has never seen observed. Regardless of other factors, machine learning algorithms consist of the following:

- **Representation :** It is the landscape of possible models, basically the space of allowed models, i.e. the hypothesis space. It is about expressing models in some formal language. For example, SVM (Support Vector Machines) with RBF kernels form one type of representation, and 3-layer feed forward neural networks (or computational graphs) form another.

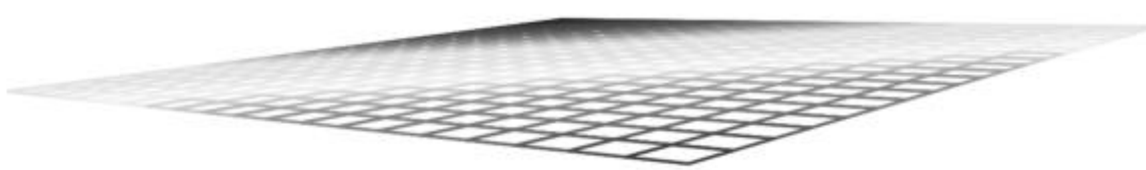

Fig 4.1 : Representation: The Model Landscape [29]

- **Evaluation :** It is judging one model vs. another. It can be considered as height of the landscape for each given model, with lower areas being more desirable than higher areas. It can also be seen as loss function, utility function, fitness function, or scoring function. For example, Mean squared error, or likelihood, or the estimated probability of a model given the observed data imply different heights at each point on a single landscape.

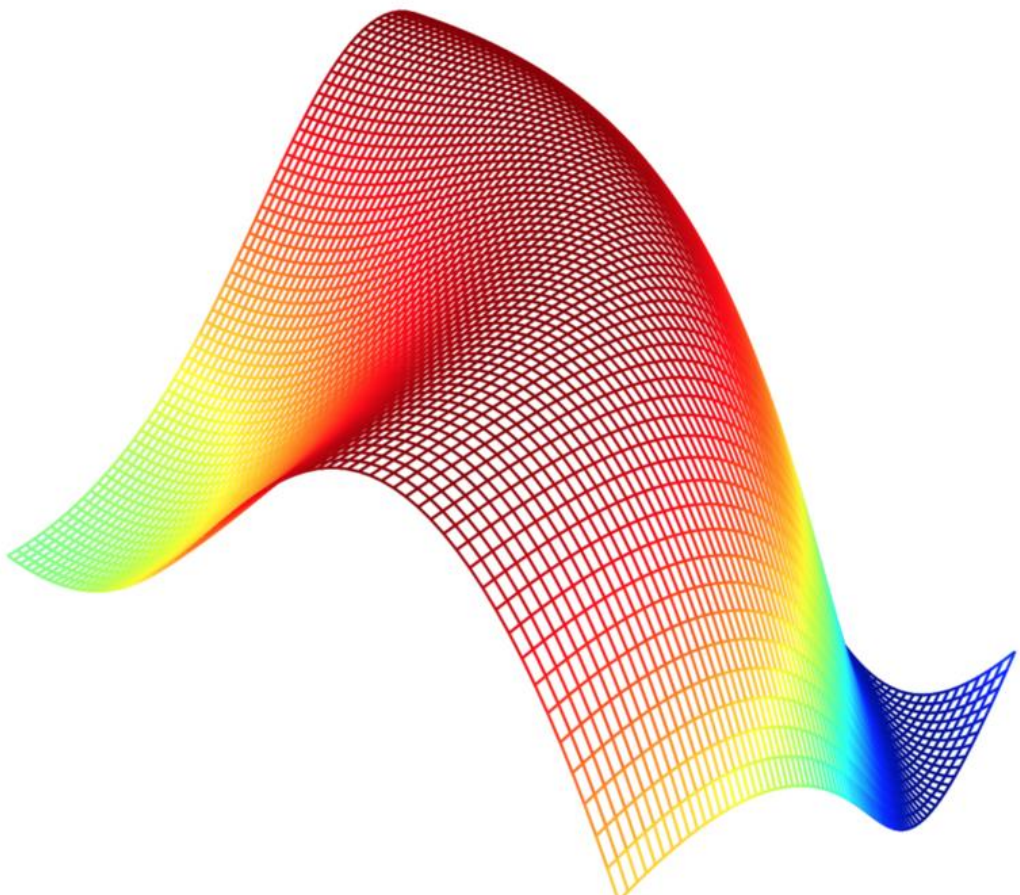

Fig 4.2 : Evaluation: Preferences over the landscape [29]

- **Optimization :** It is to obtain better evaluations, by searching the space of represented models. It can be seen as the way to traverse the landscape to find the promised land of ideal models. For example, Genetic algorithms and Stochastic gradient are two different methods of optimizing a model class.

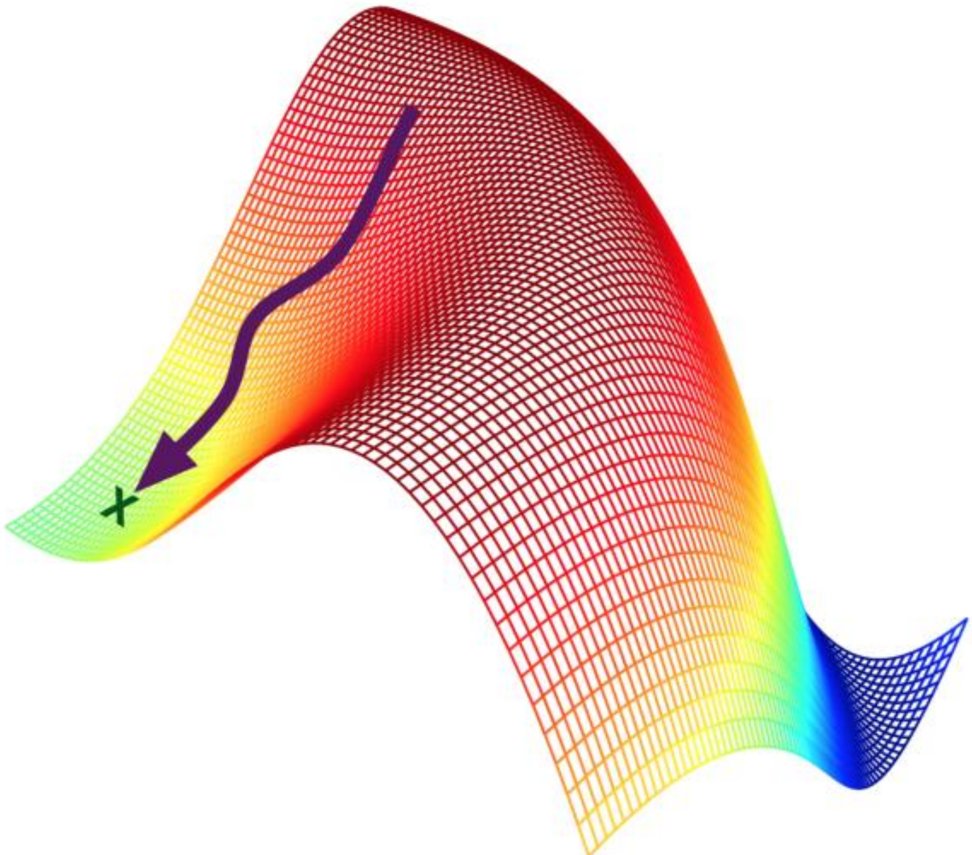

Fig 4.3 : Optimization: Strategy for fulfilling preferences [29]

Three fundamental research basis of machine learning are:

- **Task-oriented studies :** Also known as engineering approach, it is analysis as well as development of learning frameworks to enhance output in a pre-determined set of tasks.
- **Cognitive simulation :** The exploration and computer simulation of human intelligence processes.
- **Theoretical analysis :** It is the abstract exploration of the space of viable learning techniques and algorithms that are autonomous of application area.

4.1.1 Types of Machine Learning

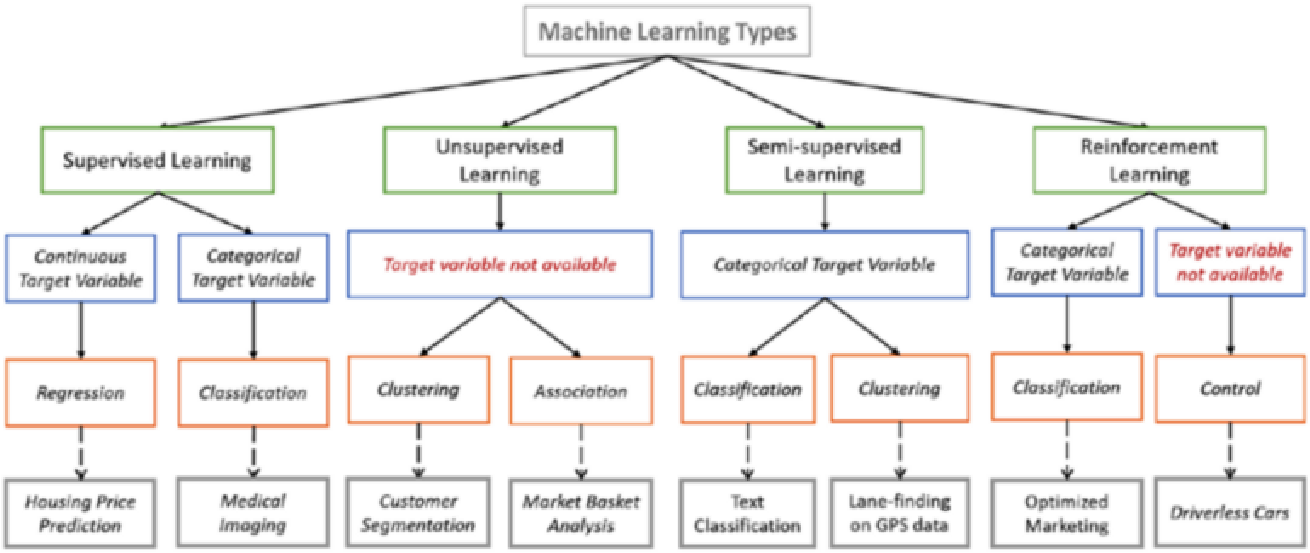

Fig 4.4 : Types of Machine Learning [30]

- **Supervised Learning :** In this type of machine learning, in order to predict successfully the response, an algorithm learns from related target reactions and example data that consist of string names or numeric values, for example classes and tags. The algorithm yields a function that maps inputs to wanted outputs. For example, X is input and Y is yield, a calculation is utilized to take in the processing capacity called yield with the assistance of this information.

$$Y = f(X)$$

It usually leaves the likelihood for inputs unspecified, and is used in techniques that are based on the information provided by the pre-determined classifications. It breaks down the information for preparation and inherently leads to the production of a derived capacity, which can be used for the mapping of new cases or instances.

Supervised Learning Algorithms:

- Decision Trees
- Naive Bayes
- Linear Regression
- Neural Networks
- Nearest Neighbour
- Support Vector Machines (SVM)

Supervised Learning Examples:

- Voice and Speech recognition
- Classification of SMS/e-mails as spam or ham
- Labelling of web pages on the basis of their content

Classification and Regression problems is the further categorization of Supervised Learning:

- **Classification :** Here, the output variable is a category, such as "yes" or "no" or some other category. Results are predicted in a discrete output, or, in other words, input variables are mapped into discrete categories. For example, Given EEG data of brain signals, to predict whether eye is in open state or closed state.
- **Regression :** Here, the output variable is a real value, such as "weight" or "height" etc. Results are predicted as a continuous output, or input variables are mapped to some continuous function. For example, to predict average price of a house given some attributes.

➢ **Unsupervised Learning :** In this type of machine learning, there is only input data and no corresponding output variables, on which the algorithm can try to model relationships. The primary objective for this type is to model underlying distribution or structure in the data for learning more about it by detecting patterns, mining for rules, and summarizing and grouping the data points. Here, learner is trained on the set of data with unlabeled instances.

These kind of machine learning algorithms are used in pattern detection and descriptive modelling.

It can be further classified into clustering and association problem:

- **Clustering :** In this problem, inherent groupings in the data are discovered, or arrangement of instances into subsets or sub-populations so that the instances which are similar to each other or are comparable and belong to same cluster or group, such as grouping customers according to purchasing behaviour, or gathering clients by obtaining conduct.
- **Association :** In this problem, rules that explain large portions of the data are expounded, such as people that buy A also tend to buy B.

Unsupervised Learning Algorithms:

- For clustering problems
  - K means algorithm
- For association rule learning problems
  - Apriori algorithm

- **Semi-supervised Learning :** In this type of machine learning, there is huge amount of input data (X) and just a portion of the data is labelled (Y). Since it can be time-consuming or expensive to label data as it requires domain experts to do this, numerous real world machine learning problems come under this area, and these algorithms are the best candidates for model building.

  Semi-supervised Learning Algorithms:

  - Generative Models
  - SVMs (Support Vector Machines)
  - Graph-Based Algorithms

- **Reinforcement Learning :** These types of machine learning algorithms continuously learn from the environment in iterative manner. In this, software agents determine the ideal behaviour within a specific context, so that its performance is maximized. For agent, to learn its behaviour, reward feedback is required, also known as the reinforcement signal.

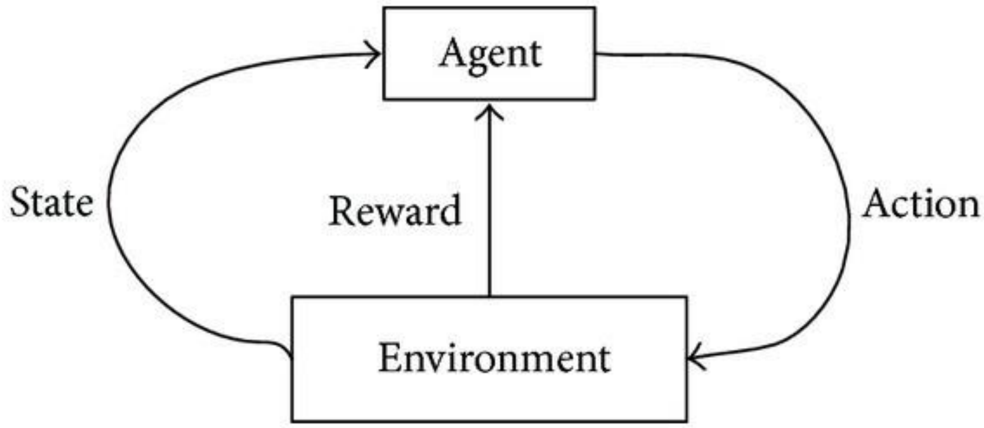

Fig 4.5 : Reinforcement Learning [31]

Reinforcement Learning Algorithms:

- Q-Learning
- Deep Adversarial Networks
- Temporal Difference (TD)

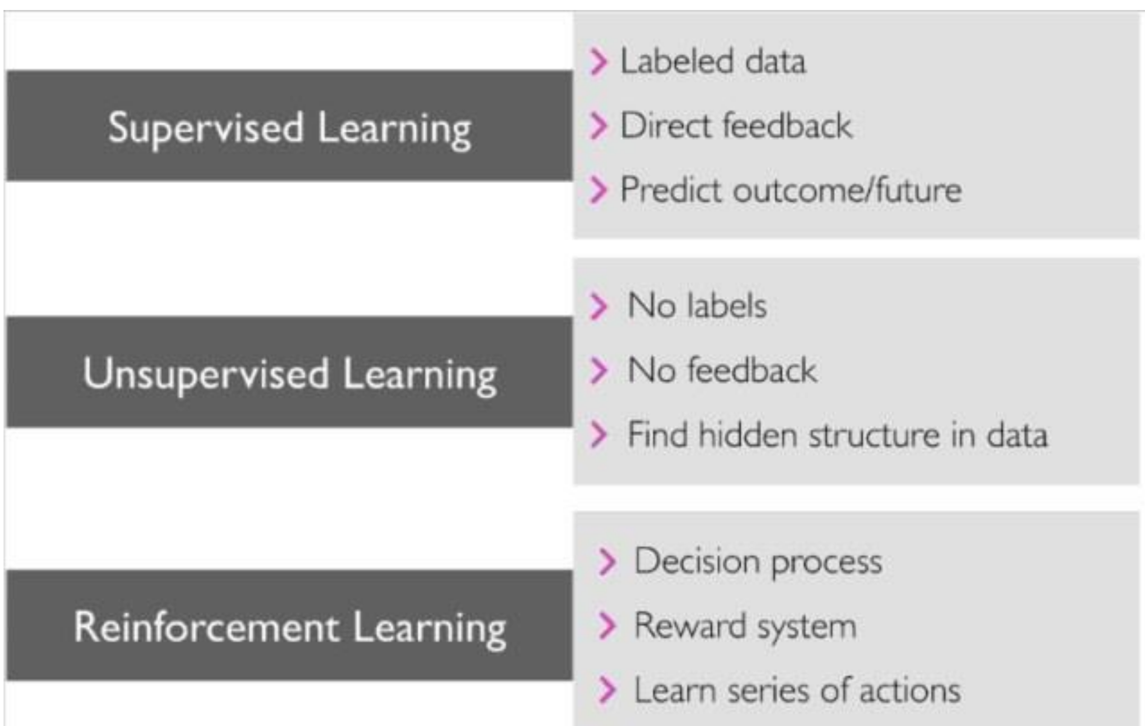

Fig 4.6 : Features of Machine Learning Types [32]

4.2 Dataset Used

The objective of this dataset is to detect eye state from EEG (Electroencephalography) data. It is available free on UCI repository. This dataset contains 14 attributes having real values, i.e AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4.

F- Frontal

T- Temporal

P- Pareital

O- Occupital

FC- Between F and C

AF- Intermediate between F_P and F

During the EEG measurement, detection of eye state was done with a camera and later, after analysis of the video frames, it was added manually to the file. The whole data is from one EEG measurement with Emotiv EEG Neuroheadset, and 117 seconds was the duration of measurement. '0' indicates the eye-open state, and '1' indicates the eye-closed state. It contains 14980 instances, and 14 variables with no missing values. There are 6723 instances of class 1 and 8257 instances of class 0.

4.3 Proposed Approach

The proposed technique compares performance of different machine learning models and then combines top 5 models on the basis of accuracy to generate ensemble model with maximum accuracy.

Since dataset used is binary class classification dataset, majority voting technique is used for ensembling. In this, final class label is predicted that has been predicted most frequently by the classification models.

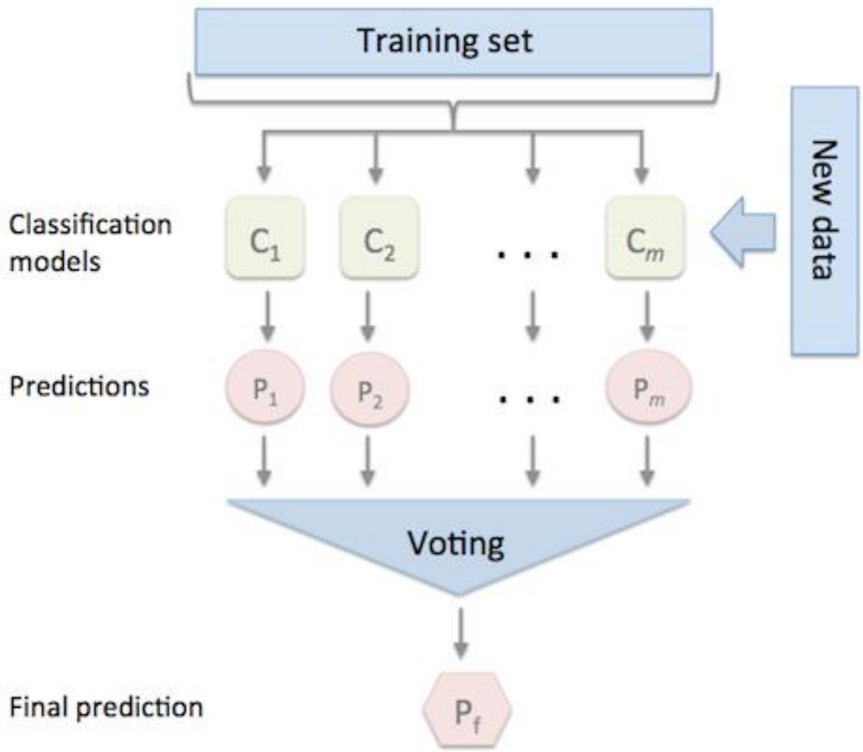

Fig 4.7 : Majority voting ensemble technique [33]

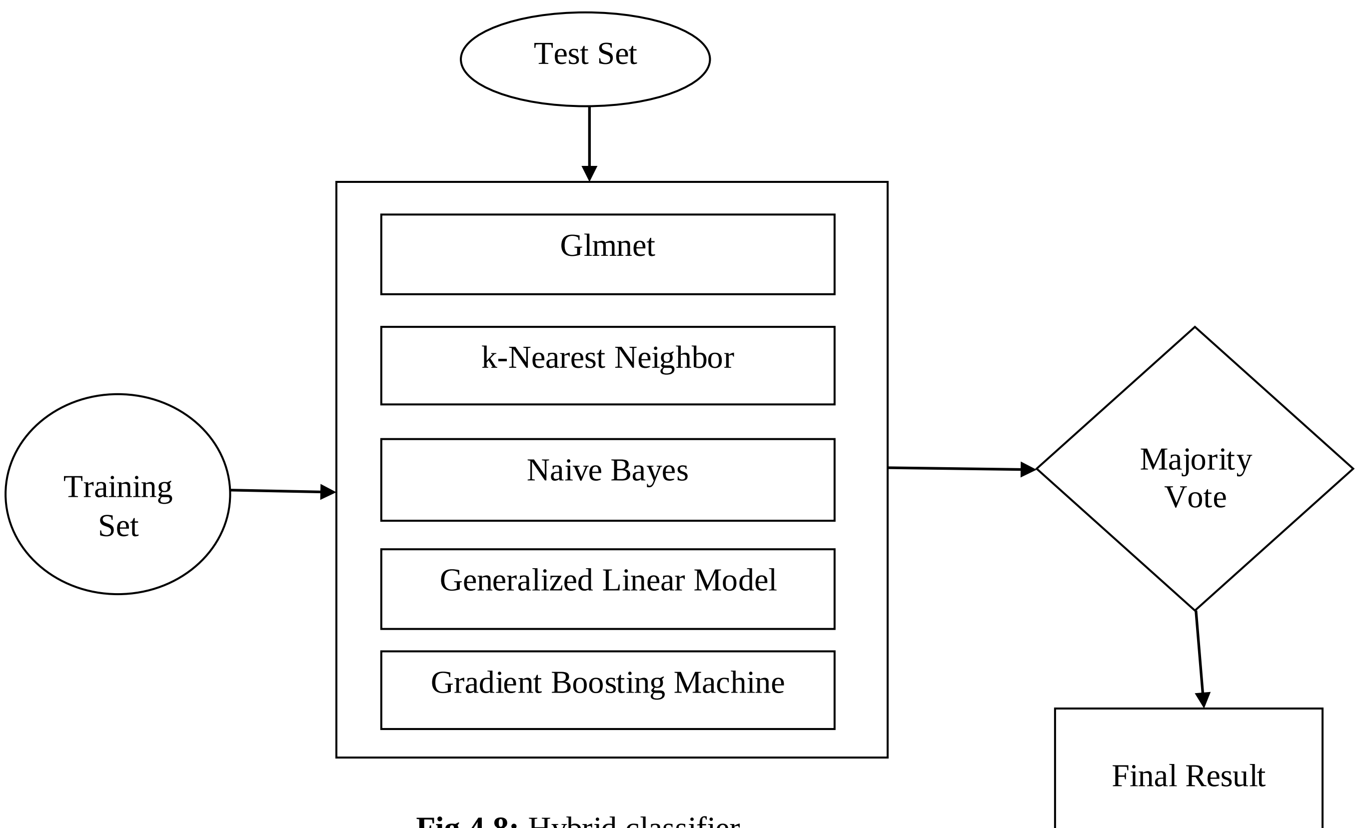

Fig 4.8: Hybrid classifier

The models used are explained as follows:

4.3.1 Decision Tree

In decision trees, on the basis of a specific parameter, data is continuously split. It is one of the types of Supervised Machine Learning algorithms. A decision tree is drawn upside down with its root at the top. The leaves are the final outcomes or decisions, and decision nodes are where data is split. Each decision node corresponds to one of the input variables, or is labeled with an input feature. The values possessed by leaf nodes are predefined classes on the dataset. Features at upper level of the tree are most important features, and features with less importance are at lower level of tree or not present in the tree. Decision tree is built upon the divide-and-conquer technique.

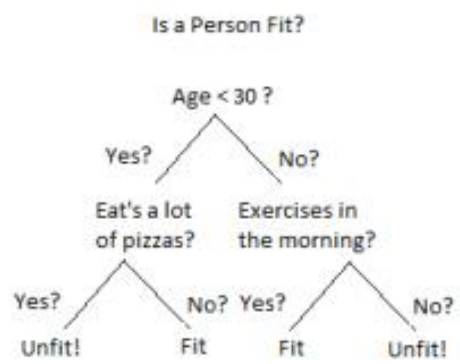

Fig 4.9 : Decision Tree [34]

In above example, the goal is to make decision whether a person is fit or unfit. This is done by making an informed decision on the basis of age, whether the person eats a lot of pizzas, and whether the person exercises in the morning. For example, if a person is of age more than 30 years and exercises in the morning, then he/she is fit.

This model has a simplicity as relations can be viewed easily and feature importance is clear. Generally, these are referred to as Classification and Regression Trees or CART.

Classification Trees: In this kind of Decision Tree, outcome variable is like Yes/No or discrete set of values. Decision variable here is Categorical.

Regression Trees: In this, outcome variable is Continuous, such as any number.

By splitting source set into subsets based on an attribute values, decision tree can learn. This process is repeated on each subset recursively known as recursive partitioning, or ID3(Iterative Dichotomiser 3) Algorithm.

The process of building Decision Tree starts by selecting a root node having close relationship with the output variable. Then, further features are selected as internal or decision nodes calculating entropy or information gain.

Entropy: Also known as Shannon Entropy, it is measure of randomness or uncertainty in data, and is denoted by H(S) for a finite set S.

$$H(S) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)} \tag{1}$$

It gives information about the predictability of a certain event. If value of entropy is 0, means the sample is completely homogeneous, and the sample is equally divided if it has entropy 1. Lower values imply less uncertainty and higher values imply high uncertainty, such as if all members belong to the same class, entropy is 0, and when half members belong to one class and other half belong to other class, entropy is 1.

Information Gain: Denoted by IG(S,A), for a set S, also known as Kullback-Leibler divergence, it is measure of effective change in entropy after deciding on a particular attribute A.

$$IG(S, A) = H(S) - H(S, A) \tag{2}$$

Basically, it is the measure of relative change in entropy with respect to independent variables.

$$IG(S, A) = H(S) - \sum_{i=0}^{n} P(x) * H(x) \tag{3}$$

where, IG(S,A) is information gain after deciding on attribute A,

H(S) is entropy of the entire set,

P(x) is the probability of event x.

The node with highest value of Information Gain is chosen as parent node, and the iteration of process is done till the leaf node and completion of decision tree.

4.3.2 Adaboost

Adaboost, stands for Adaptive Boosting, is an ensemble classifier like random forest. Ensemble classifiers are the ones that are made up of multiple classifier algorithms and their output is combined result of output of other algorithms. Its main goal is to get a strong classifier from a set of weak classifiers.

$$F(x) = sign(\sum_{m=1}^{M} \theta_m f_m(x)), \tag{4}$$

where f_m is m^{th} weak classifier,

θ_m is the corresponding weight.

Adaboost is weighted combination of m weak classifiers. In this algorithm, every instance is weighted in the training dataset, and the initial weight is set as:

$$weight(x_i)=1/n$$

where n is the number of training instances,

x_i is i^{th} training instance

Steps of adaboost for m iterations:

i. Weak classifiers are fit into the dataset and the one with lowest weighted classification error is selected.

$$\epsilon_m = E_{w_m}[1_{y \neq f(x)}] \tag{5}$$

ii. Weight for m^{th} weak classifier is calculated.

$$\theta_m = \frac{1}{2} ln(\frac{1-\epsilon_m}{\epsilon_m}). \tag{6}$$

iii. For each data point, weight is updated.

$$w_{m+1}(x_i, y_i) = \frac{w_m(x_i, y_i) exp[-\theta_m y_i f_m(x_i)]}{Z_m}, \tag{7}$$

where Z_m is normalization factor.

After all iterations, final prediction is get by summing up the weighted prediction of all classifiers

4.3.3 Generalized Linear Model

Mean in the traditional regression models was dependent on the explanatory variables by a link function, and Generalized Linear Model is its extension, or in other words, it is a versatile generalization of ordinary linear regression, that have error distribution models instead of a normal distribution. It is unification of other statistical models for example Poisson, logistic, and linear regression.

It contains 3 components:

i. **A random component:** It indicates the conditional distribution of response variable, Y_i for the explanatory variables values in the model.

ii. **A linear predictor:** It is in relation with the expected value of data via link function, and in the model, incorporates information about the independent variables.

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

where η denotes a linear predictor, and is expressed as linear combinations of unknown parameters, β and matrix of independent variables, X.

iii. **A smooth link function:** Distribution function mean and linear predictor are related through this link function.

It differs from general linear model in 2 ways:

i. Linear combination of predictor variables are used to predict the dependent variable values, which are connected via a link function.
ii. The distribution of the response or dependent variable does not have to be continuous. It can be non-normal.

4.3.4 SVM (Support Vector Machine)

It is a supervised machine learning algorithm that can be used for both classification and regression problems, but mainly used for purpose of binary classification. Here, every data item is represented as a point in n-dimensional space, and then classification is performed by expounding the hyperplane by which the margin between the two classes is maximized. Support vectors are the cases that elucidate the hyperplane.

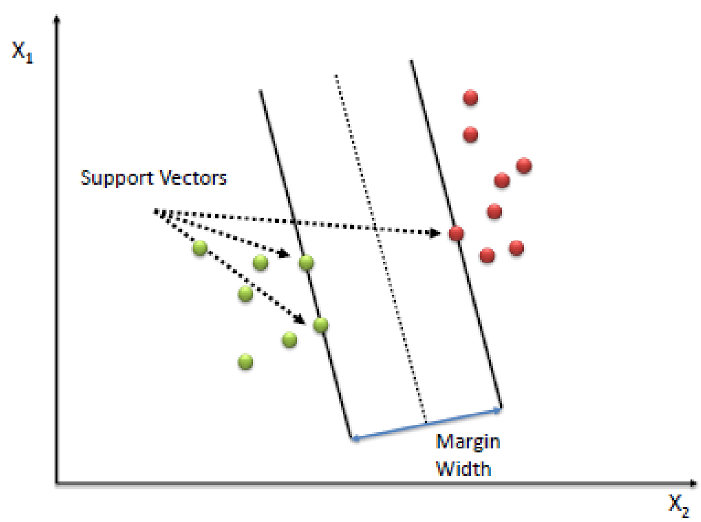

Fig 4.10 : Support Vector Machine [35]

The beauty of this algorithm lies in the fact that there is unique global minimum value if the data is linearly separable. Ideally, SVM should generate a hyperplane through which the vectors are completely separated into two non-overlapping classes, but this may not be possible practically, and hence a hyperplane is found that minimizes the misclassifications and maximizes the margin. SVM uses iterative training algorithm for constructing an optimal hyperplane so that error function is minimized. New output is predicted using the dot product between each support vector (x_i) and the input (x) as:

$$f(x) = B_0 + \mathrm{sum}(a_i * (x,x_i)) \tag{8}$$

where the coefficients a_i and B_0 are evaluated from the training data through learning algorithm. Implementation of SVM is done through a kernel. Number of kernel functions are used in Support Vector Machine models such as Radial Basis Function (RBF), linear, polynomial, and sigmoid.

$$K(\mathbf{X}_i, \mathbf{X}_j) = \begin{Bmatrix} \mathbf{X}_i \cdot \mathbf{X}_j & \text{Linear} \\ (\gamma \mathbf{X}_i \cdot \mathbf{X}_j + \mathrm{C})^d & \text{Polynomial} \\ \exp(-\gamma \mid \mathbf{X}_i - \mathbf{X}_j \mid^2) & \text{RBF} \\ \tanh(\gamma \mathbf{X}_i \cdot \mathbf{X}_j + \mathrm{C}) & \text{Sigmoid} \end{Bmatrix} \tag{9}$$

$$K(\mathbf{X}_i, \mathbf{X}_j) = \phi(\mathbf{X}_i) \bullet \phi(\mathbf{X}_j) \tag{10}$$

For some specific kernel functions, gamma is an adjustable parameter. Dot product of input data points with higher dimensional feature space by transformation ϕ is represented as kernel function.

SVM has wide application in biological and other sciences. They can also be used in classification of images, image segmentation systems, and text and hypertext categorization.

4.3.5 Random Forest

In this algorithm, multiple decision trees are build up and then merged to get a more stable and accurate prediction.

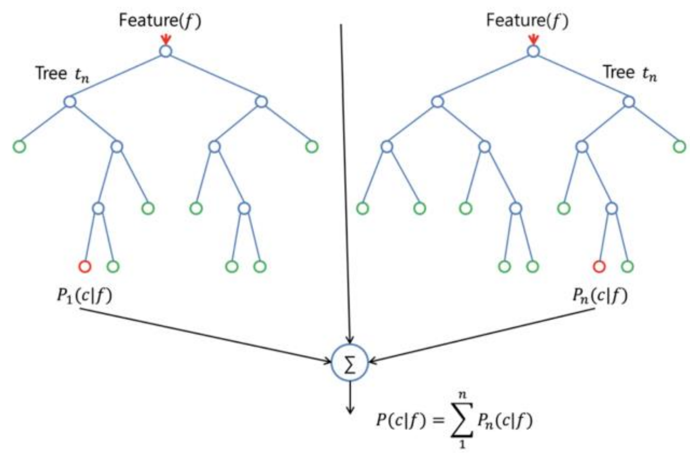

Fig 4.11 : Random Forest [36]

It is basically an ensemble learning method with fundamental that a "strong learner" can be formed from a group of "weak learners". Random Forests attempt to mitigate problems of high bias and high variance which are present in single decision trees. Instead of searching for the best feature during node split, it hunts for the best feature among a random subset of features, hence brings extra randomness in the model. This process results in a better model.

Correlation and strength of every individual tree in the forest are the two factors on which forest error rate depends. More correlation implies more forest error rate. The one with low error rate is a strong classifier (tree), and increasing strength of individual trees decreases the forest error rate.

Some of the advantages of random forests include:

- Reduces overfitting
- Widely used
- No need for feature normalization
- Individual Decision Trees can be trained in parallel, therefore parallelizable

It also has some disadvantages:

- Not easily interpretable
- Not a state-of-the-art method

4.3.6 GBM (Gradient Boosting Machine)

In this algorithm, model is build up in a stage-wise manner as done by other boosting algorithms and generalized with optimization of an arbitrary differentiable loss function. This method works in following manner:

i. Optimize loss function.
ii. Predictions are made through a weak learner.
iii. Loss function is minimized by an additive model which adds weak learners.

Boosting is an ensemble method in which predictors are made sequentially and not independently, and Gradient Boosting Machine is an example of boosting algorithm.

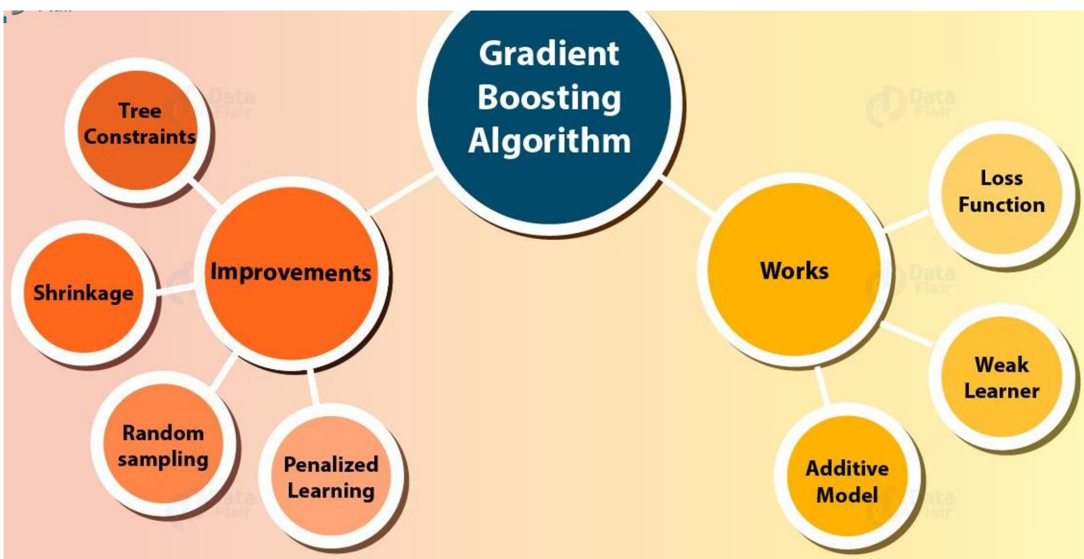

Fig 4.12 : Gradient Boosting Algorithm [37]

4.3.7 KNN (k-Nearest Neighbors)

k-Nearest Neighbors is an algorithm in which all available cases are stored and a case is classified by majority vote of its neighbors. A neighbor can be determined using different methods of distance, such as Euclidean and Hamming distance etc. It is a type of lazy learning.

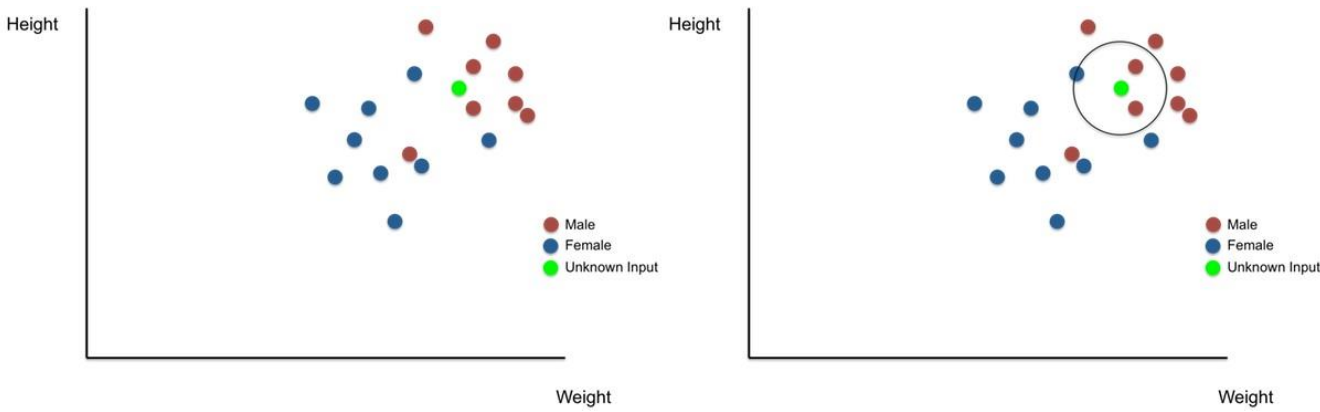

Fig 4.13 : k-Nearest Neighbors [38]

For some new instance, x, predictions are made by searching through entire training set for the k most similar neighbors or instances, and then for those k neighbors, the output variable is summarized. kNN is used both for classification and regression. For classification, the output is most common class value, and for regression, it is mean output variable.

Distance functions

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \qquad (11)$$

Fig 4.14 : Distance functions used in k-Nearest Neighbors

Some advantages of kNN include its versatility, high accuracy, simplicity, and no assumptions about data, and is useful for non linear data.

Some disadvantages of kNN include that it is computationally expensive since it stores all of the training data, high memory requirement, slow prediction in case of big N, and sensitive to irrelevant features and scale of the data.

4.3.8 Naive Bayes

Naive Bayes classifier belongs to family of "probabilistic classifiers", and is a classification algorithm used for binary and multi-class classification problems. Naive Bayes classifier uses following kinds of probabilities:

- **Conditional Probabilities:** The conditional probabilities of each input value given each class variable.
- **Class Probabilities:** The probabilities of each class in the training dataset.

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c) \tag{12}$$

Naive Bayes model is simple and easy to build, hence it is used for very large data sets, and outperform highly sophisticated classification methods.

Advantages of Naive Bayes:

- In case of categorical input variables, it performs very well. It assumes normal distribution in case of numerical variables.

- It is fast and easy to predict class of test data set, and performs very well in multi class prediction.
- Less training data is required in this classifier.

Disadvantages of Naive Bayes:

- It assumes predictors to be independent, which is impossible in real life problems.
- "Zero Frequency" is another limitation of Naive Bayes. "Zero Frequency" is the case when some variable has a category, which is not there in training set, model will assign 0 probability and will not be able to make a decision. To solve this problem, Laplace estimation is used.
- It is a bad estimator.

4.3.9 Glmnet

This model fits generalized linear model through maximum likelihood. It fits multinomial, logistic, poisson, Cox, and linear models, and can also fit multi-response linear regression. It can exploit sparsity in the input matrix, and is very fast.

Chapter 5

Results

For the comparison of various classifiers used on the dataset, parameters used are:

- Accuracy
- Sensitivity
- Specificity

For the calculation of above parameters, confusion matrix is used.

Confusion Matrix : It is used to calculate the performance of a classifier on a set of test data for which true values are known.

		ACTUAL CLASS	
PREDICTED CLASS		0	1
	0	True Positive (TP)	False Positive (FP)
	1	True Negative (TN)	False Negative (FN)

Table 5.1 : Confusion Matrix

True Positive : The instances that actually belong to class 0 and are correctly classified to class 0.

True Negative : The instances that actually belong to class 1 and are correctly classified to class 1.

False Negative : The instances that actually belong to class 0 and are incorrectly classified to class 1.

False Positive : The instances that actually belong to class 1 and are incorrectly classified to class 0.

Accuracy : It is the fraction of predictions that the applied classifier got correct. In other words, it is a vital parameter for calculating performance of any classifier that how well that classifier predicts the target value in test dataset as compared to its actual value. More accuracy implies more better model. Accuracy is calculated using the formula:

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

Accuracy of classifiers used in eye state prediction is:

CLASSIFIER	ACCURACY (%)
Decision Tree	72.04
Adaboost	79.78
Support Vector Machine	84.87
Random Forest	92.39
Gradient Boosting Machine	97.69
Generalized Linear Model	97.69
Naive Bayes	97.82
k-Nearest Neighbor	97.84
Glmnet	97.98

Hybrid	98.73

Table 5.2 : Accuracy of various classifiers

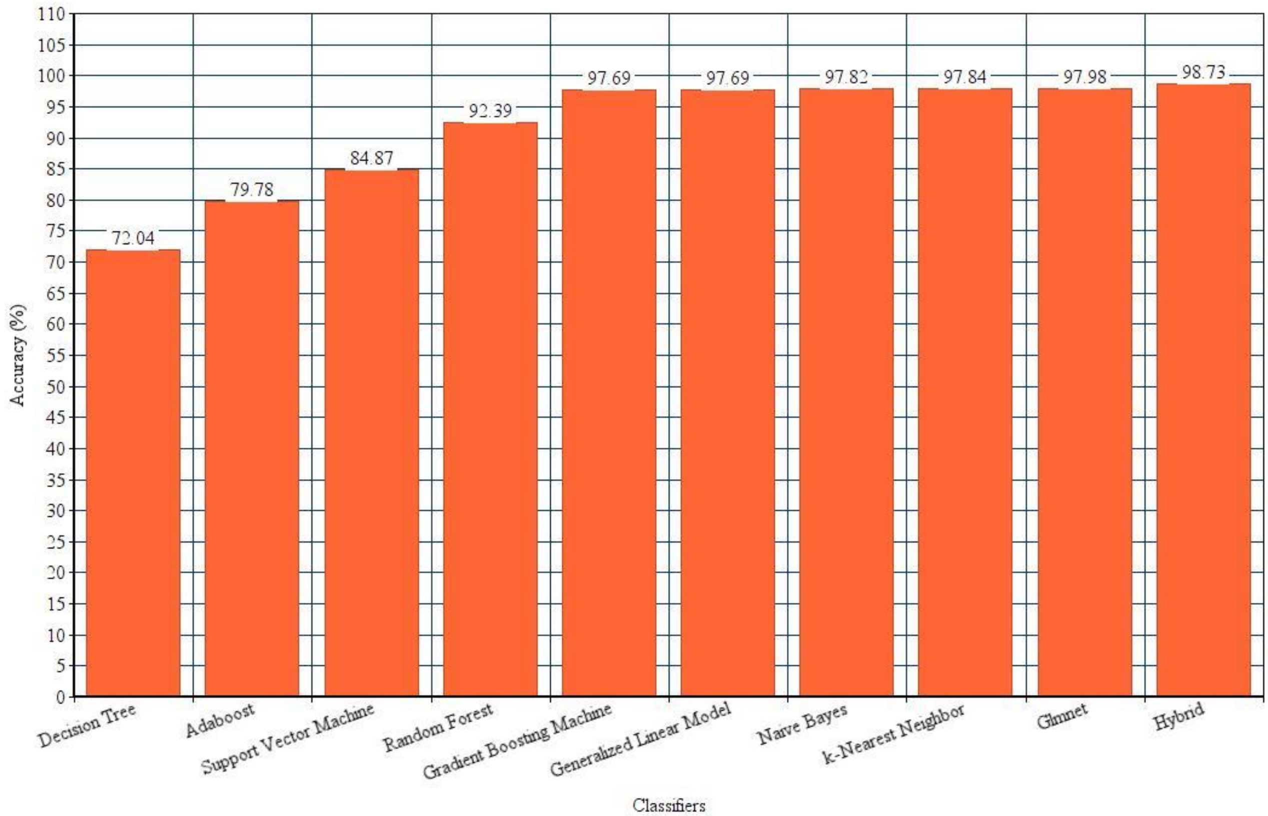

Fig 5.1 : Graph of Accuracy of various classifiers

Sensitivity : Also known as True Positive Rate, it refers to the proportion of actual positives that are correctly identified. It is calculated by the formula:

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

Sensitivity of classifiers used in eye state prediction is:

CLASSIFIER	**SENSITIVITY**
Decision Tree	0.559
Adaboost	0.74
Support Vector Machine	0.747
Random Forest	0.895
Gradient Boosting Machine	0.968
Generalized Linear Model	0.966
Naive Bayes	0.97
k-Nearest Neighbor	0.971
Glmnet	0.97
Hybrid	0.976

Table 5.3 : Sensitivity of various classifiers

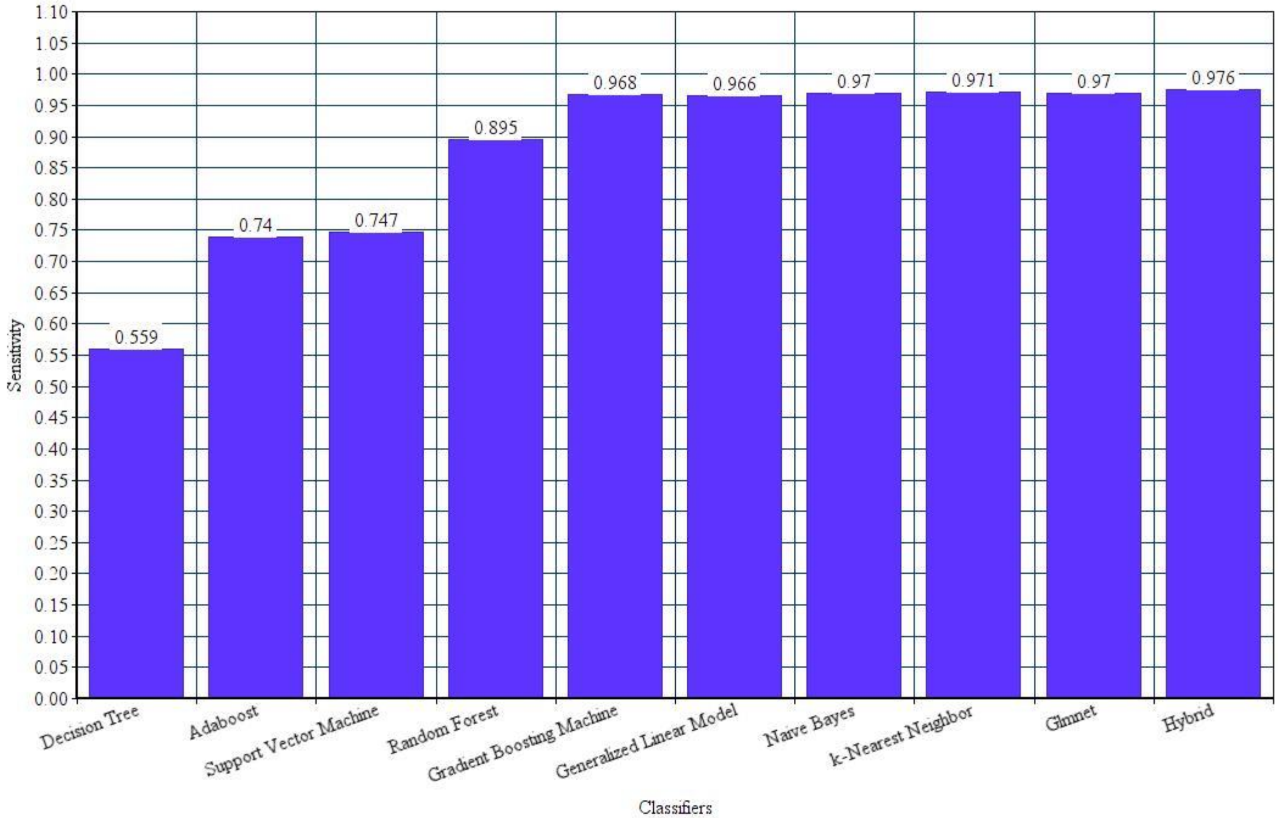

Fig 5.2 : Graph of Sensitivity of various classifiers

Specificity : Also known as True Negative Rate, it refers to the proportion of actual negatives that are correctly identified. It is calculated by the formula:

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

Specificity of classifiers used in eye state prediction is:

CLASSIFIER	SPECIFICITY
Decision Tree	0.852

Adaboost	0.844
Support Vector Machine	0.931
Random Forest	0.947
Gradient Boosting Machine	0.984
Generalized Linear Model	0.986
Naive Bayes	0.985
k-Nearest Neighbor	0.984
Glmnet	0.988
Hybrid	0.991

Table 5.4 : Specificity of various classifiers

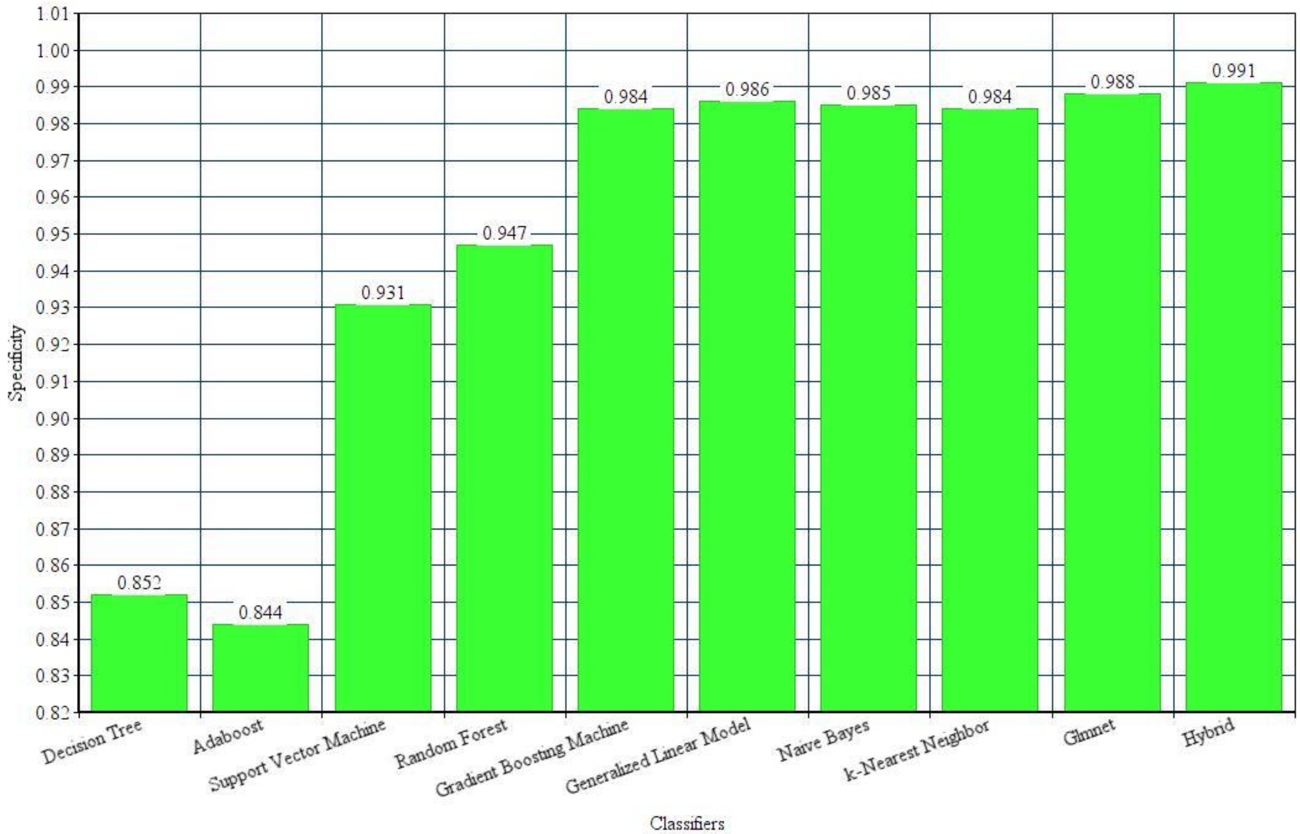

Fig 5.3 : Graph of Specificity of various classifiers

Chapter 6

Conclusion and Future Scope

This thesis presented a hybrid ensemble model for eye state prediction with accuracy of 98.73%, which is acceptable for various practical applications. The proposed technique is much more accountable, robust, and scalable than earlier used techniques. More accurate will be prediction of eye state, more accurately other activities such as driver drowsiness, stress, etc. can be predicted, and will be beneficial in many ways.

6.1 Future Scope

Research is an iterative and continuous process. Future scope of this field will definitely pave a way for researchers.

- Here, EEG data was used for eye state prediction. In future, other recording methods such as Electro-corticography (ECoG) can be used for eye state prediction.
- In this, supervised learning techniques were used for classification. In future, unsupervised and semi-supervised learning techniques can be applied and relevant features can be extracted, thus increasing understanding of brain activity.
- It would be interesting to see whether feature selection reduces number of sensors without any compromise in the performance, as cost will reduce with reduction in number of sensors and speed also will increase.
- This model be further enhanced and expanded to improve accuracy further and using minimum features.

References

[1] Ghiyasvand, M. S., et al. "A new EEG signal processing technique for discrimination of eyes close and eyes open." *Engineering in Medicine and Biology Society, 1995 and 14th Conference of the Biomedical Engineering Society of India. An International Meeting, Proceedings of the First Regional Conference., IEEE*. IEEE, 1995.

[2] Vachiravel, Sabarinathan. (2015). Eye State Prediction using EEG Signal and C4.5 Decision tree algorithm. International Journal of Applied Engineering Research. 10. 167-171.

[3] Hamilton, Cameron R., Shervin Shahryari, and Khaled M. Rasheed. "Eye State Prediction from EEG Data Using Boosted Rotational Forests." *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*. IEEE, 2015.

[4] Rösler, Oliver, and David Suendermann. "A first step towards eye state prediction using eeg." *Proc. of the AIHLS* (2013).

[5] Yeo, Mervyn VM, et al. "Can SVM be used for automatic EEG detection of drowsiness during car driving?." *Safety Science*47.1 (2009): 115-124.

[6] Chambayil, Brijil, Rajesh Singla, and Rameshwar Jha. "Virtual keyboard BCI using Eye blinks in EEG." *Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on*. IEEE, 2010.

[7] Wolpaw, Jonathan R., and Dennis J. McFarland. "Multichannel EEG-based brain-computer communication." *Electroencephalography and clinical Neurophysiology* 90.6 (1994): 444-449.

[8] Jacob, Robert JK. "What you look at is what you get: eye movement-based interaction techniques." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1990.

[9] Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research, vol. 3, 2003.*

[10] S. Ganesh Kumar, V. V. Ramalingam, and V. Sugumaran, ”Analysis of EEG signals using data mining approach,” *IJCET, vol. 3, pp. 206-212, June 2012.*

[11] La Rocca, Daria, Patrizio Campisi, and Gaetano Scarano. "EEG biometrics for individual recognition in resting state with closed eyes." *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*. IEEE, 2012.

[12] Li, Ling, Lei Xiao, and Long Chen. "Differences of EEG between eyes-open and eyes-closed states based on autoregressive method." *Journal of Electronic Science and Technology* 7.2 (2009): 175-179.

[13] O. Ossmy, O. Tam, R. Puzis, L. Rokach, O. Inbar, and Y. Elovici, "Mind Desktop - Computer Accessibility for Severely Handicapped,” *Proc. of the ICEIS, Beijing, China, 2011.*

[14] Gopan, Gopika, Neelam Sinha, and Dinesh Babu. "Statistical feature analysis for EEG baseline classification: Eyes Open vs Eyes Closed." *Region 10 Conference (TENCON), 2016 IEEE*. IEEE, 2016.

[15] Li-Hong, Zhao, Duan Ning, and Yang Cai-Kun. "Fatigue-driving recognition based on the state of the human eye and electroencephalographic signal." *Control and Decision Conference (CCDC), 2016 Chinese*. IEEE, 2016.

[16] Reddy, Tharun Kumar, and Laxmidhar Behera. "Online Eye state recognition from EEG data using Deep architectures." *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016.

[17] Thuraisingham, Ranjit A., et al. "Analysis of eyes open, eye closed EEG signals using second-order difference plot." *Medical & biological engineering & computing* 45.12 (2007): 1243-1249.

[18] Thuraisingham, Ranjit A., et al. "Frequency analysis of eyes open and eyes closed EEG signals using the Hilbert-Huang transform." *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012.

[19] Wang, Ting, et al. "Time series classification for EEG eye state identification based on incremental attribute learning." *Computer, Consumer and Control (IS3C), 2014 International Symposium on*. IEEE, 2014.

[20] Trejo, Leonard J., Roman Rosipal, and Bryan Matthews. "Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials." *IEEE transactions on neural systems and rehabilitation engineering* 14.2 (2006): 225-229.

[21] Sadatnezhad, Khadijeh, Reza Boostani, and Ahmad Ghanizadeh. "Classification of BMD and ADHD patients using their EEG signals." *Expert Systems with Applications* 38.3 (2011): 1956-1963.

[22] Neha Jain, Sandeep Bhargava, Savita Shivani, and Dinesh Goyal. Eye State Prediction using EEG by Supervised Learning. International Journal of Science, Engineering and Technology. 10.2348/ijset06150715

[23] Fitriyani, Norma Latif, Chuan-Kai Yang, and Muhammad Syafrudin. "Real-time eye state detection system using haar cascade classifier and circular hough transform." *Consumer Electronics, 2016 IEEE 5th Global Conference on*. IEEE, 2016.

[24] Fuangkaew, Supakit, and Karn Patanukhom. "Eye state detection and eye sequence classification for paralyzed patient interaction." *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*. IEEE, 2013.

[25] Xu, Cui, Ying Zheng, and Zengfu Wang. "Efficient eye states detection in real-time for drowsy driving monitoring system." *Information and Automation, 2008. ICIA 2008. International Conference on*. IEEE, 2008.

[26] https://searchsqlserver.techtarget.com/definition/data-mining

[27] http://www.datascienceassn.org/content/data-mining-process

[28] http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html

[29] https://medium.com/@devnag/machine-learning-representation-evaluation-optimization-fc7b26b38fdb

[30] https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

[31] https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

[32] https://www.kdnuggets.com/2017/11/3-different-types-machine-learning.html

[33] https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/

[34] https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html

[35] http://www.saedsayad.com/support_vector_machine.htm

[36] https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

[37] https://data-flair.training/blogs/gradient-boosting-algorithm/

[38] https://brilliant.org/wiki/k-nearest-neighbors/

ABC

ORIGINALITY REPORT

11%	2%	7%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Tharun Kumar Reddy, Laxmidhar Behera. "Online Eye state recognition from EEG data using Deep architectures", 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016 Publication	1%
2	"Progress in Computing, Analytics and Networking", Springer Nature, 2018 Publication	1%
3	Cameron R. Hamilton, Shervin Shahryari, Khaled M. Rasheed. "Eye State Prediction from EEG Data Using Boosted Rotational Forests", 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 2015 Publication	1%
4	Lecture Notes in Computer Science, 2012. Publication	1%
5	Submitted to Chandigarh University Student Paper	<1%

6	Submitted to Colorado Technical University Online Student Paper	<1%
7	Submitted to University of Warwick Student Paper	<1%
8	Boran, Fatih Emre, Diyar Akay, and Ronald R. Yager. "An overview of methods for linguistic summarization with fuzzy sets", Expert Systems with Applications, 2016. Publication	<1%
9	Submitted to Siddaganga Institute of Technology Student Paper	<1%
10	Fawagreh, Khaled, Mohamed Medhat Gaber, and Eyad Elyan. "Random forests: from early developments to recent advancements", Systems Science & Control Engineering, 2014. Publication	<1%
11	Submitted to International University - VNUHCM Student Paper	<1%
12	blog.hackerearth.com Internet Source	<1%
13	Florin Leon. "Machine Learning Methods Used to Predict the Liquid-Crystalline Behavior of Some Copolyethers", Molecular Crystals and	<1%

Liquid Crystals, 1/2007
Publication

14	Submitted to University of Wales Swansea Student Paper	<1%
15	Submitted to ABV-Indian Institute of Information Technology and Management Gwalior Student Paper	<1%
16	Submitted to Indian School of Business Student Paper	<1%
17	Submitted to University of Dayton Student Paper	<1%
18	Submitted to VIT University Student Paper	<1%
19	Submitted to Nottingham Trent University Student Paper	<1%
20	www.comp.nus.edu.sg Internet Source	<1%
21	Submitted to University of Edinburgh Student Paper	<1%
22	Submitted to University of Maryland, University College Student Paper	<1%
23	Submitted to University of Dallas Student Paper	<1%

24	Submitted to University of Newcastle upon Tyne Student Paper	<1%
25	Anita Wasilewska. "A Model PM for Preprocessing and Data Mining Proper Process", Lecture Notes in Computer Science, 2007 Publication	<1%
26	Submitted to University of New York in Tirana Student Paper	<1%
27	Gokulakrishnan, Balakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Ragavan, Nadarajah Prasath, and AShehan Perera. "Opinion mining and sentiment analysis on a Twitter data stream", International Conference on Advances in ICT for Emerging Regions (ICTer2012), 2012. Publication	<1%
28	Gerard G. Dumancas, Ghalib A. Bello, Jeff Hughes, Renita Murimi et al. "chapter 30 Visualization Tools for Big Data Analytics in Quantitative Chemical Analysis", IGI Global, 2018 Publication	<1%
29	Bülent Yılmaz, Sümeyye Korkmaz, Dilek Betül Arslan, Evrim Güngör, Musa H. Asyalı.	<1%

"Like/dislike analysis using EEG: Determination of most discriminative channels and frequencies", Computer Methods and Programs in Biomedicine, 2014
Publication

---

30 Selvakumar, K., Jovitha Jerome, Kumar Rajamani, and Nishanth Shankar. "Real-Time Vision Based Driver Drowsiness Detection Using Partial Least Squares Analysis", Journal of Signal Processing Systems, 2015.
Publication
<1%

---

31 www.j3.jstage.jst.go.jp
Internet Source
<1%

---

32 Submitted to Universiti Teknologi Petronas
Student Paper
<1%

---

33 www.frontiersin.org
Internet Source
<1%

---

34 www.sans.org
Internet Source
<1%

---

35 Armin Lawi, M Sya'Rani Machrizzandi. "Facial Expression Recognition using Multiclass Ensemble Least-Square Support Vector Machine", Journal of Physics: Conference Series, 2018
Publication
<1%

---

36 www.dezyre.com

Internet Source

<1%

Exclude quotes	Off	Exclude matches	< 8 words
Exclude bibliography	On		