Eye State Prediction Using Ensembled Machine Learning Models

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

Master of Engineering

in

Information Security

Submitted by

Dipali Singla

(801433010)

Under the Supervision of

Dr. Prashant Singh Rana

Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

June 2016

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Eye State Prediction using Ensembled Machine Learning Models*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in Information Security in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Prashant Singh Rana 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.

(Dipali Singla)

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

(Dr. Prashant Singh Rana)

Assistant Professor
Computer Science and Engineering Department
Thapar University

Countersigned by

(Dr. Maninder Singh)

Head,CSED

Thapar University

Patiala

(Dr. S.S. Bhatia)

Dean (Academic Affairs)

Thapar University

Patiala

ACKNOWLEDGEMENT

Words are often too less to reveals one's deep regards. An understanding of the work like this is never the outcome of the efforts of a single person. I take this opportunity to express my profound sense of gratitude and respect to all those who helped me through the duration of thesis.

This work would not have been possible without the encouragement and able guidance of my supervisor, Dr. Prashant Singh Rana, Assistant Professor, CSED, Thapar University, Patiala. Most of the novel ideas and solutions found in this thesis are the result of our numerous stimulating discussions. Their feedback and editorial comments were also invaluable for the writing of this thesis.

I would also like to thank all the faculty members of the department and my friends who directly or indirectly helped me in completion of my thesis. No words of thanks are enough for my dear parents and brother whose support and care makes me stay on earth. Thanks to be with me.

Above all, I would like to thank the Almighty God for his blessings and for driving me with faith, hope and courage in thinnest of the times.

Place: Thapar University, Patiala

Dipali Singla

Abstract

As electric signals are transmitted between the brain cells for transferring data within the brain, capturing of these signals can result in understanding the functionality of brain and other directly linked parts (like eyes, ears, spinal nerves etc) of our body. It is done by Electro Encephalogram Test (EEG). Along with capturing Normal electric signals we can also capture epileptic seizures which are caused due to disruption in the normal working of brain. These electric signals are to be captured by small electrodes placed on human scalp using a standard 10/20 system on an Electro Encephalograph monitor in form of waves. These wave forms are transmitted to form of data for getting required information from data collected. In this dissertation, we will predict the state of eye (open or closed) by exploring 13 machine learning models on a 15 features dataset of an EEG test. The records of 14 electrodes are used for this prediction. Machine learning models in R language are statistical analysis and prediction analysis methods used on dataset by training and testing of the dataset. Results are evaluated using 6 different machine learning parameters i.e. Sensitivity, Confusion matrix, Kappa value, Specificity, Accuracy and Receiver Operating Characteristics (ROC) curve. K- Fold validation and assembling of models will be done on best three predictive models pertaining to our dataset.

Table of Contests

List of Figures

List of Tables

Table No. **Title** **Page No.**

Chapter 1

Introduction

An EEG (Electro Encephalogram) test is to analyze activities of brain. All the data transfer between the brain cells occur through the electric impulse signals. The wave pattern formed from these signals are recorded and analyzed to detect the proper functioning of brain. These wave patterns are obtained by electrodes attached to the scalp of brain[1, 2]. Now, if any sort of problem occurs in some part of the brain, it can also be detected by analyzing the inappropriate or irregular wave signals formed by the electrodes attached to that part of scalp. All electrodes are attached to the scalp, according to the standard 10/20 international system, as shown in Fig 1. This system tells all fixed points and nasion-inion distance between the electrodes. EEG captures information about the brain cells and can also be used to analyze some other information from wave signals (eg. Eyes information, nervous system information attachment to the brain, sleepiness of a person or presence of foreign objects in our brain etc). Also, a newly trending use of EEG system is to create Brain Computer Interface(BCI) systems. These systems can be used to develop an interface for people with severe motor disabilities, so that they can communicate conveniently with other people and are capable to control the virtual environment.

In this work, we are using some of the electrodes information collected from the EEG test to predict the state of an eye. Electrodes used in this work are O1, O2, FC6, AF3, F8, AF4, F4, FC5, T7, T8, F3, P7, F7, P8. The placement of these electrodes can be seen in Fig 1. Even numbered electrodes are placed on right hemisphere of scalp, whereas odd numbered are placed on the left hemisphere of scalp. The 15th features of the dataset represents eye state in binary form(0 or 1). 0 represents that the eye is open whereas 1 stands for closed eye state. Machine learning in R is a process of data collection, feature selection, training and testing with predictive models, finally giving output in the form of prediction value, accuracy and some other performance parameters. These models are then compared on the basis of these parameters to find the best suitable models for the existing dataset. 13 machine learning models are explored on this dataset. All the models are described in Table 4.1. Then, models are ensembled for obtaining better efficiency in terms of result parameters i.e. accuracy, sensitivity, specificity, ROC curve and confusion matrix. Following chapters describes the parameters and models used in this research work.

1.1 10/20 System

The 10/20 system is worldwide accepted technique to define the position of scalp electrodes. The electroencephalogram's (EEG) wave forms are detected using electrodes which are placed on scalp. the recordings are stored on the machine called an electroencephalograph. All the brain activity is captured using these signals. Thousands of brain cells called as neurons are simultaneously involved in the transmission of these signals. As the arousal of the person changes, the activity in these wave form also changes. Signal is low when the person is in relaxed state. while they are in high frequency if the person is very excited. These recorded wave forms can be used to determine a number of purposes like sleep research, in diagnosis of any kind of epilepsy. Positioning of the EEG electrodes is done by a single technique which is used internationally. In this technique, 10 and 20 stands for distance. This distance is between the two adjoining electrodes, which has to be either 10 or 20 per cent of the whole skull of human on which test has to be implemented.

There are total 21 electrodes positioned on the surface of the scalp. Each area has an alphabet to recognize the node and also a numeral to recognize the hemisphere position. There is no central node in it. 'C' symbol is used only for the purpose of recognition only. 'Z' symbol relates to an electrode which is placed on the mind line. The alphabet F stands for 'Frontal Lobe', T stands for 'Temporal Lobe', C stands for 'Central Lobe', P stands for 'Parietal Lobe' and O stands for 'Occipital Lobe'. Even numerals (2, 4, 6, and 8) describe the location of electrode on the right hemisphere. Odd numerals (1, 3, 5, and 7) describe the location of electrode on the left hemisphere. Four anatomical indicators are used for the crucial positioning of electrodes: First indicator is the nasion which is a mark between brow and snout, second indicator is the inion which is the lowest mark of the skull from the back of the head and is normally represented by a distinguished bump and the pre detectable marks antecedent to the ear.

1.2 Brain Waves

Brain waves medically termed as neural oscillations. Brain waves are different wave patterns formed from the electric signals used by neurons to communicate with each

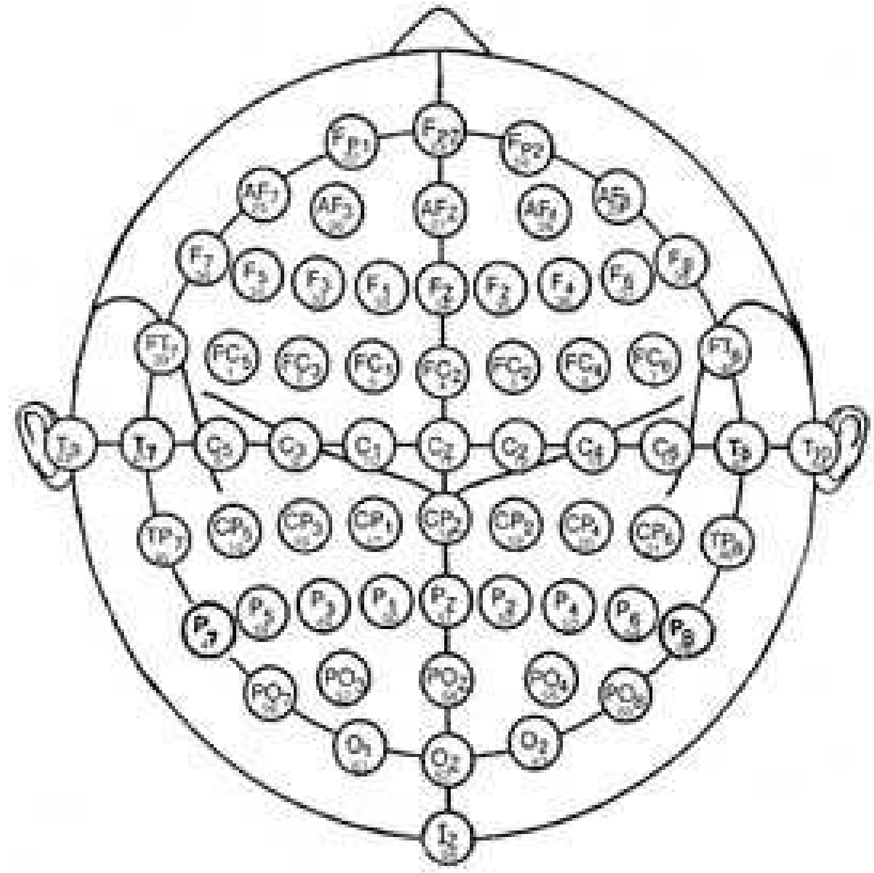

Figure 1.1: 10/20 system for EEG test.

other. Each brain wave has a definite meaning to transmit which helps serve us to know optimal mental functioning These are captured by electrode sensors placed on the scalp during the EEG Test. With the change in our mood, feelings and activeness level pattern and frequency of wave forms also do change. Higher the frequency more active and more crazily we are thinking or feeling. Brain waves are measured in hertz (Hz). Different types of wave forms captured are listed below:

1.2.1 Delta Waves

These are the slowest but loud wave form recorded. These wave forms have high wave frequency. They are mostly found in infants and young children. With the increasing age frequency in creation of these wave form also decreases. This wave form depict deepest levels of relaxation and restorative, healing sleep, unconsciousness. Delta waves suspend external awareness and are the source of empathy. These waves are generated when human body is healing itself or resettling internal clock signals. Delta waves: They never go down to zero because that would mean that your brain is dead. But, deep dreamless sleep would take you down to the lowest frequency. Typically, 2 to 3 cycles a second. A person dont dream in this wave form. Major key features and diagram **??** of delta wave are given below:

- Ranges between 0.5 - 3 Hz.
- Slowest wave form.
- Generated when the person is in deep sleep or unconscious.
- One dont dream in this wave form.
- They have also been found to be involved in unconscious bodily functions such as regulating heart beat and digestion.
- If delta waves are not normal, then a person can suffer from learning disability or difficulty in maintaining mental awareness.
- These are also generated when a person is depressed or feeling low.

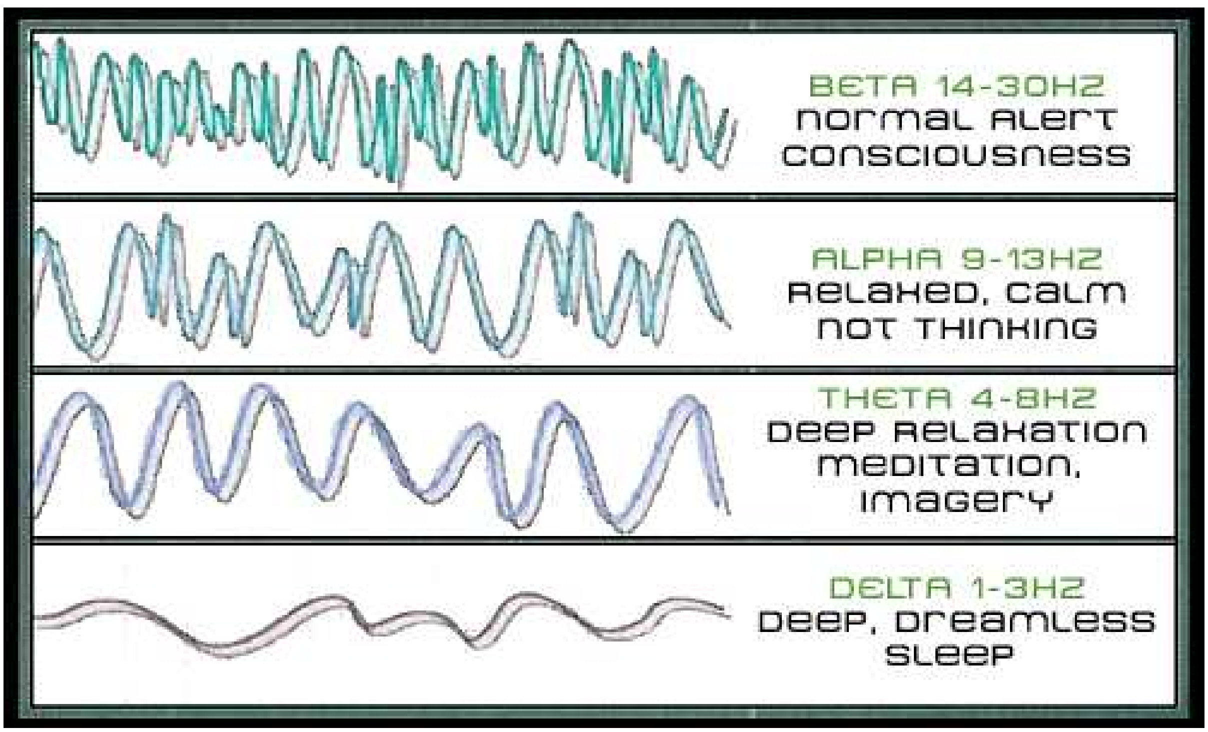

Figure 1.2: Different wave frequencies

1.2.2 Theta wave

Theta brain waves are considered brain waves that oscillate between the frequencies of 4 Hz to 8 Hz. These waves are mostly found in children and adults suffering from Attention Deficit Hyperactivity Disorder(ADHD) or the individuals having facing difficulty in concentrating on a particular task. These waves are transmitted majorly when the individual is feeling emotional, intuition, daydreaming, relaxation, REM sleep and the subconscious mind. It can also be said these waves are transmitted when person is in subconscious thinking, spiritual experiences or tied to supernatural. Most of activities or experience that took place during this time cannot be remembered by the individual, as they are in deep sleep or totally unconscious. Theta waves in some regions can be due to person facing some disability in learning. In this we hold our fears, history or other dramatic or tragic stuff.

- Children haves more theta activity than adults.
- It is associated with deep states of relaxation and sleep.
- A very little theta activity is found in adults when they are awake.
- These are mostly transmitted when the person is unconscious or during deep sleep, hence they are not able to remember any experience.
- Presence of theta waves makes concentration extremely difficult.
- It is a state of somnolence with reduced consciousness.
- It is proven to be useful in hypnotherapy.
- These are dominant in meditation.
- During the dominance of these waves are either in dream or in institution or in a imaginary world which is beyond the conscious awareness.
- Theta activity in our brain is increased due to some of the following factors:
    1. Attention Deficit Hyperactivity Disorder(ADHD)
    2. Deep sleep
    3. Emotional
    4. Hyperfocus
    5. Impulsitivity
    6. Meditation/ Spriritual activity
    7. Long-time/ childhood memories
    8. Subconscious mind

1.2.3 Alpha wave

These wave form ranges from 8 Hz to 12 Hz. These wave forms are originated/transmitted when person in physically or mentally relaxing but is well aware of activities going in surrounding, e.g. when you just get up in the morning after sleep you are naturally in that state. It is mostly created in right hemisphere of brain. It is also connected to when one recalls old memory, lessen pain or reduce stress. Alpha is resting state of mind. It helps in calmness, alertness, learning, mind/Body integration and most importantly overall mental coordination. Alpha activity is reduced as soon as eyes are opened. It is very important part as it calms our brain and refreshes from any mental stress or strain. It is a state in which a person is neither fully awake nor fully sleep, just relaxes along with acknowledging the activities taking place in the environment.

In occipital lobe, these waves increases in high amount when the person is in relaxing state while keeping the eyes closed. The person is not asleep but mentally relaxed, keeping thought process on a lower level. In REM sleep, waves are formed in frontalcentral region due to our rapid eye movement. Actual or pin-pointed purpose is still not known to the researchers. It indicates slightly aroused. In slow-wave these waves are found during the alpha-delta state. Some researchers believe that they indicate awaking state during deep sleep.

- Mostly found in right hemisphere of brain.
- These exist in relaxed but awake state, i.e. person is not processing anything.
- Research shows that experienced mediators have strong increase in alpha activity.
- These are considered slower wave form in brain wave forms.
- These wave form stand for stand for Power of Now.
- Alpha activity is reduced as soon as eyes are opened.
- It calms our brain and thus refreshes it.
- It is state between sleep and waking state.
- In slow-wave these waves are found during the alpha-delta state.
- Alpha Waves are mostly found due to :
    1. Calmness
    2. Creativity

3. Balance mood
4. Day dreaming
5. Flow state of mind
6. Positive thinking
7. Relaxation
8. Relaxation

1.2.4 Beta waves

Beta waves oscillates over a wide range of 12 Hz to 40 Hz. These occur in people which are fully alert and focused. These are emitted when one is focused on external stimuli or putting some extra mental effort. These brain waves are formed when person is in state of alertness, energy, concentration, tense, afraid or focused. People suffering from ADHD have very low degree of beta waves detection. It is present mostly during all day long waking hours. Lacking of sufficient beta activity can lead to a number of disorders like mental and emotional depression or insomnia. These waves can be divided in three bands:

Low Beta Waves - (12 Hz 15 Hz) These are also known as Beta 1 waves. The lower range of Beta activity is often associated mostly with quiet, focused, introverted concentration. These are highly thought as fast idle waves.

Mid-Range Beta Waves (15 Hz 20 Hz) These are also known as Beta 2 waves. It is linked with increase in physical and mental activity in terms of energy, performance, anxiety or when the person is involved in deep thought process or in state of excitement while having new experiences.

High Beta Waves (18 Hz 40 Hz) These are also known as Beta 3. These are high oscillating waves transmitted when high level of stress, paranoia, high arousal, highly excited or anxiety. These are transited to gamma waves. It contains high level of energy level.

Some of factors responsible for generation of beta waves are:

- People which are hypothesized are detected with high beta beta activity in wrong parts of brain.
- It is very useful in sensory feedback and motor control.

- People suffering from ADHD have very low degree of beta waves detection.
- Lacking of sufficient beta activity can lead to a number of disorders like mental and emotional depression or insomnia.
- These are transmitted when the person is engaged in judgment, problem solving, decision making, highly emotional agitation or focused in mental activity.
- Transmission of these waves will make a person to have high focus, good memory, beta function like spot on and is able to concentrate better.
- Too much of beta activity can result in insomnia or negativity.
- Beta Waves are mostly found in:
  1. Alertness
  2. Attention
  3. Awake
  4. Anger
  5. Energy
  6. Excitement
  7. IQ increase

1.2.5 Gamma waves

These waves are highly oscillating waves ranges over 60 hz. These have very high frequency just as flute. These are associated when there is formation of ideas or memory processing or various types of learning. These waves disappear when the person is in deep sleep. These transmit information very fastly. It is highly active when brain is in state of altruism, high virtue, spiritually connected or in universal love. These waves are associated with increased problem in self-control, intelligence, compassion and problem solving. Brain injury can cause to shortage of gamma waves.

- These disappear when the person is in deep sleep.
- These waves are transited from high beta waves.
- How these are formed is still a mystery.

- These waves are associated with increased problem in self-control, memorization, intelligence, compassion and problem solving.
- Shortage of these waves can cause learning disability, cognitive functioning and mental retardation.
- Gamma Waves are mostly found due to:
  1. Attention
  2. Mental processing
  3. Unity of consciousness
  4. Meditation
  5. REM sleep.

Chapter 2

Literature Survey

Rosler et al [1], investigated the prediction of eye state by measuring brain waves with EEG test. A corpus is recorded containing the activation strength of fourteen electrodes recorded by a EEG Emotiv EPOC headset and manually stating the eye state corresponding the data. 42 Machine Learning algorithms are performed through open-source software Weka to predict the eye state from the trained corpus. Other researchers can easily reproduce these obtained data. Classification error rate are results to be achieved from this investigation. Applications of this study will be in obtaining brain stimuli as an input mode for computer games, to obtain emotions and then store and analyze to create a virtual Robotic environment, for handicapped persons to control their surrounding electronic control devices and can also be used in military scenarios. The prediction is done with accuracy more than 97 per cent. High accuracy is obtained inspite of the fact that no special training is given to the algorithms. Further a number other dependency factors can also considered by which eye state differs or gets affected.

Yeo et al [3] performs pattern recognition is done by a superior signal classification tool i.e. Support Vector Machines (SVM). SVM is useful in identifying and differentiating EEG changes between alert state and drowsy state of driver, which is the main objective of this paper. This paper proposes a technology similar to ANN to classify and predict Drowsy state. Driving simulation with EEG monitoring id done on twenty people. Alert State is known by the presence of Dominant beta wave, while alpha dropouts indicate drowsy state. Fast and slow eye blinks also show the alertness of the driver. The recorded data of alert and drowsy state is used to train SVM program. SVM model predicts transition between two states i.e. either from alert to drowsy state or from drowsy state to alert state. Many EEG characteristic features along with wave form correlation is studied to differentiate between drowsy and alert state. The general agreement for reliable detection of fatigue and drowsiness is the visually inspection of EEG Waves. EEG signals are analyzed either by Spectral Pattern Recognition or by signal pattern recognition. Manual classification is done after the checks are applied by trained SVM program on unclassified data set. A preemptive automatic drowsy detection system can be formed for security purposes.

Fukuda et al [4], propose a pattern classification method to evaluate whether EEG signals can be used as human interface tool, which is measured by a simple electroence phalograph. Signals are recorded for 450 minutes by exposing eyes to torch light (switching between on or off) and using this records for classification. A log-linearized Gaussian mixture Neural network along with a stat model are explored on record data. Classification done on training data gives a varied rate of change in classification depending upon size of training subset. Feature selection on the basis of feature vectors gives the best results. Pattern recognition is done using neural networks incorporated with Pdf model is proposed to improve a generalized ability to get the expected high classification performance . Standard deviation and classification rates are pictured to understand the effects on classification results by LLGMN on the eye state data. Basic classification is done is done by LLGMN in four layers with sufficient accuracy. Feed forward Neural network are explored to improve convergence rate and classification rate. Feature selection adds on to the process of improving efficiency.

Sahu (2015) et al [5] In this paper binary classification is performed on the EEG dataset to categorize state of eye. Classification is similar to data mining in which find patterns in similar groups. Dataset feature selection is a pre-processing step in machine learning. Incremental Feature Reordering is a subset which is obtained by the feature selection performed on the dataset. And gives all the non-dominant features for EEG signal corpus to create a reorder set. By removing all the non-dominant features optimal subset is obtained thus increasing the efficiency and accuracy. Feature ordering uses correlation function as basic parameter to rearrange the dataset feature and obtain optimal subset. Double linked list is used data structure to create dynamic environment for reordering the featured set.

S. Natarajan et al. [6] proposed an approach for classification and detection of tuberculosis. It is a disease due to mycobacterium which attacks low immune bodies and spreads through air. Proposed methodology was the combination of classification and clustering that divide tuberculosis into two parts retroviral and pulmonary tuberculosis. In this paper they have used k-means clustering which divide tuberculosis data into 2 clusters and defined classes for each cluster. K-means clustering was combined with classification algorithms to improve performance and accuracy for the prediction of tuberculosis. Subsequently various classifiers was trained on tuberculosis data to build classifier model using k-fold cross validation approach. Here, proposed approach helps for diagnosis decisions and their treatments.

Jesse Davis et al. [7] presented relation between receiver operator characteristic and precision- recall curve. ROC curve used for binary decision problems and PR curve give information about performance of classification algorithms in machine learning. Researcher was show that curve dominates in precisionrecall space it will also dominate in receiver operator characteristic curve. In this paper researcher was also show a method for computing precision-recall curve and receiver operator characteristic curve.

Xue-wen Chen et al. [8] proposed feature selection approach, feature assessment by sliding threshold, that measure importance of features with the help of area under the ROC in one dimensional feature space using sliding decision line. By using Roc curve or rank features they have created another issues i.e. where to place threshold. Possible solution was to use histogram to find where threshold was placed.

Touraj Varaee et al. [9] presented hybrid variable selection approach which uses concept of wrapper technique along with lower cost and also improve performance of classifiers. Proposed method was the combination of sample domain filtering, two feature selection method and re-sampling for refining of sample domain. Here, approach was divided into two phase, in first phase filters was used and in second phase they have used the concept of wrapper subset selection and genetic search. First phase refine and analyze the sample domain for better result in second phase. In second phase filtering technique eliminates irrelevant features and wrapper method selects relevant features with higher accuracy and lower cost.

Taghi M. Khoshgoftaar et al. [10] this paper includes comparative study of bagging and boosting techniques for noisy binary class data and imbalanced data. When data are clean but imbalance, boosting and bagging was less significant. Bagging can be used without replacement to deal with imbalanced and noisy data.

Tian-Yu Liu et al. [11] proposed method called as mutual information based on feature selection for easy ensemble to deal with load balancing and improve performance of easy ensemble classifier and compared with support vector machine and easy ensemble. Proposed method improves prediction ability. They have used concept of mutual information to describe dependency between two random variable .It is also referred to relative entropy. Mutual information among two variables can be defined as follows: Where p m,n is probability distribution. Proposed method was train model on training set by easy ensemble and calculates mutual information on training set. After that it will select relevant features and ranking them and generate optimal training subset and retrain model on training subset.

Kehan Gao et al. [12] in this paper they have investigated a methodology of variable selection with ensemble learning process and examined 2 learning method and 5 feature selection techniques namely filter based feature ranking techniques i.e., chi squared, symmetrical uncertainty, information gain, wrapper method and embedded method.

Huan Liu et al. [13] focused on feature selection or variable selection techniques in machine learning. Variable selection is important step for the machine learning applications and it remove irrelevant and redundant features and also reduce dimension.

Chapter 3

Problem Statement

EEG test is used to capture and analyze the brain activity and also to analyze functioning of other related parts. State of eye can also be analyzed by the captured wave form patterns being transmitted between the brain cells in the EEG test. Electrodes are placed on the human scalp to capture the brain signals according to the 10/20 internationally adopted system. My objective is to train the machine learning models with EEG dataset, so that they can predict the future values of the prediction attribute i.e. state of an eye that can be obtained from the EEG wave form pattern. Ensembling and K-fold validation is to be performed on best three models

3.1 Objective

- To explore machine learning models on the dataset.
- To select top three models suited for the dataset.
- Ensemble top three models for better performance.
- Validate the ensembled model performance using k-fold validation.

Chapter 4

Dataset

4.1 Introduction

A data set of about 15000 Records with 15 features is used to evaluate the resulting predictions. This data set is available at *https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State*. First 14 features of dataset are records of different electrodes (namely O1, O2, FC6, AF3, F8, AF4, F4, FC5, T7, T8, F3, P7, F7, P8) used in the EEG test. Information about these electrodes is shown in Table 4.1. These electrodes are placed using 10/20 international system. Even numbered electrodes are placed on right hemisphere of scalp, whereas odd numbered are placed on the left hemisphere of scalp. The target feature of our dataset is binary data about state of eye, where 0 represents that eye is open and 1 represents that eye is closed. This dataset is used to explore machine learning models to predict the state of eye. Table 4.2 shows the sample dataset used in this work.

Table 4.1: Electrode details.

Electrode	**Lobe**
F	Frontal
T	Temporal
P	Pareital
O	Occupital
FC	Between F and C
AF	Intermediate between Fp and F

4.2 Dataset Feature Evaluation

Correlation is the degree to which two or more attributes or measurements on the same group of elements show a tendency to vary together. It can also be described as ”degree and type of relationship between any two or more quantities (variables) in which they vary together over a period”.Correlation can vary over the range of +1 to -1. Values close

Table 4.2: Sample dataset.

Electrode	Value 1	Value 2	Value 3	Value 4	Value 5	Average
o1	4096.92	4097.44	4096.92	4113.33	4104.1	4110.40
o2	4641.03	4638.97	4630.26	4631.28	4626.67	4616.56
FC6	4211.28	4207.69	4206.67	4228.72	4232.82	4202.46
AF3	4329.23	4324.62	4327.69	4315.9	4323.08	4321.92
F8	4635.9	4632.82	4628.72	4625.13	4628.21	4615.21
AF4	4393.85	4384.1	4389.23	4369.23	4378.46	4316.44
F4	4280.51	4279.49	4282.05	4288.21	4293.85	4279.23
FC5	4148.21	4148.72	4156.41	4116.92	4123.08	4164.95
T7	4350.26	4342.05	4336.92	4346.15	4346.15	4341.74
T8	4238.46	4226.67	4222.05	4245.64	4244.62	4231.32
F3	4289.23	4293.85	4295.38	4255.9	4265.64	4264.02
P7	4586.15	4586.67	4583.59	4624.1	4615.38	4644.02
F7	4009.23	4004.62	4006.67	4037.44	4041.03	4009.77
P8	4222.05	4210.77	4207.69	4227.18	4229.23	4218.83
Eye State	0	0	0	1	1	0.45

to +1 indicate a high-degree of positive correlation, and values close to -1 indicate a high degree of negative correlation. The following table 4.3 shows how features of our dataset are correlated with each other.

Table 4.3: Correlation between attributes.

	F4	O1	P7	T8	O2	AF4	F7	FC5	T7	FC6	F3	F8	P8	AF3
F4	1.000	0.821	0.585	0.619	0.291	0.374	-0.165	-0.153	-0.202	-0.275	-0.327	-0.756	-0.763	-0.760
o1	0.821	1.000	0.680	0.514	0.352	0.067	-0.225	-0.227	-0.114	-0.380	-0.432	-0.744	-0.746	-0.747
P7	0.585	0.680	1.000	0.667	0.636	0.124	0.066	0.177	0.381	0.0004	0.008	-0.360	-0.362	-0.363
T8	0.619	0.514	0.667	1.000	0.674	0.520	0.062	0.204	0.375	0.333	0.198	-0.174	-0.182	-0.182
o2	0.291	0.352	0.636	0.674	1.000	0.279	0.150	0.349	0.583	0.434	0.384	0.107	0.104	0.101
AF4	0.374	0.067	0.124	0.520	0.279	1.000	0.467	0.501	0.353	0.598	0.558	0.215	0.201	0.208
F7	-0.165	-0.225	0.066	0.062	0.150	0.467	1.000	0.817	0.641	0.545	0.695	0.512	0.508	0.513
FC5	-0.153	-0.227	0.177	0.204	0.349	0.501	0.817	1.000	0.790	0.678	0.836	0.598	0.594	0.598
T7	-0.202	-0.114	0.381	0.375	0.583	0.353	0.641	0.790	1.000	0.753	0.794	0.623	0.621	0.621
FC6	-0.275	-0.380	0.0004	0.333	0.434	0.598	0.545	0.678	0.753	1.000	0.882	0.796	0.789	0.790
F3	-0.327	-0.432	0.008	0.198	0.384	0.558	0.695	0.836	0.794	0.882	1.000	0.831	0.826	0.828
F8	-0.756	-0.744	-0.360	-0.174	0.107	0.215	0.512	0.598	0.623	0.796	0.831	1.000	0.999	0.999
P8	-0.763	-0.746	-0.362	-0.182	0.104	0.201	0.508	0.594	0.621	0.789	0.826	0.999	1.000	0.999
AF3	-0.760	-0.747	-0.363	-0.182	0.101	0.208	0.513	0.598	0.621	0.790	0.828	0.999	0.999	1.000

Table 4.4: Min-Max.

Attribute	min(closed)	max(closed)	avg(close)	min(open)	max(open)	avg(open
F4	4225	4281	4368	4201	4277	7002
o1	4026	4073	4167	3581	4071	4178
P7	4574	4618	4708	2768	4620	4756
T8	4174	4233	4323	4152	4229	6674
o2	4567	4616	4695	4567	4615	7264
AF4	4246	4367	4552	1366	4356	4573
F7	3905	4005	4138	3924	4013	7804
FC5	4058	4121	4214	2453	4123	4250
T7	4309	4341	4435	2089	4341	4463
FC6	4130	4204	4319	4100	4200	5170
F3	4212	4265	4367	4197	4263	5762
F8	4212	4265	4367	4197	4263	5762
P8	4147	4202	4287	4152	4200	4586
AF3	4198	4305	4445	1030	4297	4504

4.2.1 Feature Details

4.2.2 Frontal lobe

Humans have best developed and largest frontal lobe than any other organism. Front and upper area of cortex is place where it is situated. These are also called as control centre for emotions and personality of the person. Motor functioning, spontaneity, problem solving, memory, initiation, language, social and sexual behaviour and impulse control are the major functions of this lobe. Frontal lobe is very vulnerable to physical damage because they are situated in front area of the brain. Lesions are mojor threat to the brain affecting its functioning drastically. Any sort of damage can be a cause to minute or vast change in facial expressions, personality or interpretation ability of a person eg. risk or damage analysis, stimuli reaction of a person. Some major functionalities and problems related to frontal lobe are listed below:

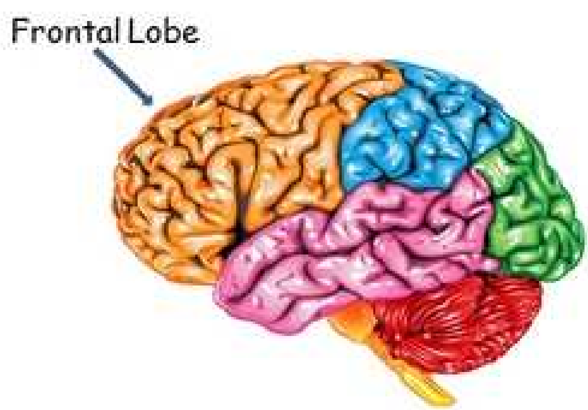

Figure 4.1: frontal lobe

Functions

- It helps in determining our behaviour (consciousness) of a person in his/her surroundings.
- Initialization ability of a person are directly dependent on the frontal lobe.
- It helps in making our judgements of routine activities.
- It is emotional control centre.
- Emotional response is highly affected, if this lobe is not working properly.
- Controls our body gestures.

- Assigns meaning to the words we choose.
- choice of correctly associated words are chosen by this lobe.
- All the motor activities and our habits are affected by this lobe.

Problems

- Loss of simple movement of various body parts (Paralysis)
- Inability to plan a sequence of complex movements needed to complete multi-stepped tasks, such as making coffee
- Loss of spontaneity in interacting with others. Loss of flexibility in thinking
- Persistence of a single thought (Perseveration)
- Inability to focus on task (Attending)
- Mood changes (Emotionally Labile)
- Changes in social behavior
- Changes in personality
- Difficulty with problem solving

4.2.3 Parietal Lobe

Parietal lobe is situated in between frontal and occipital lobe and above the temporal lobe. Primary sensory area is the main working are of the frontal lobe. This area is responsible for interpreting the impulses from the nerve cells(skin) and responding them with required stimuli. Cold, warmth, touch and pain are the major actions on which this area(lobe) reacts. Spatial information is also involved by this lobe, which gives us the ability to judge shape, size and distance. The parietal association cortex another specific area of this lobe helps us in understanding written language. It is a triangular shaped area which also helps in solving mathematical problems. Right hemisphere parietal lobe is less active than left hemisphere. Different symbols of letters and numbers are also interpreted by this lobe. It also helps in image interpretation and spatial distance.

Functions

- Location for visual attention
- Location for touch perception

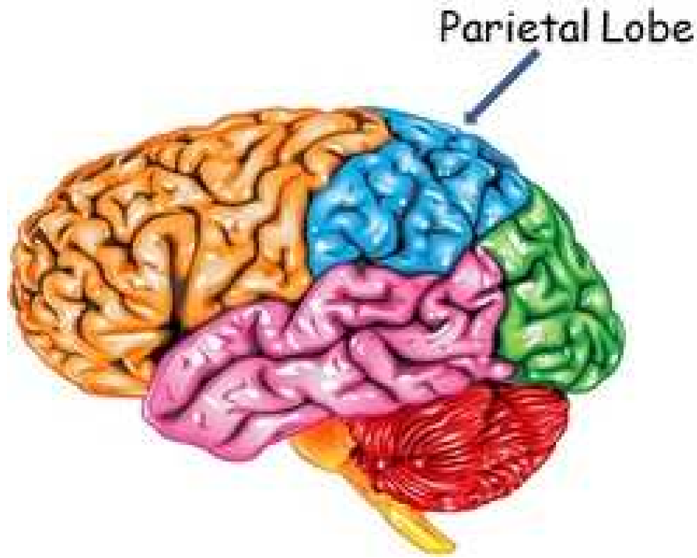

Figure 4.2: Parietal Lobe

- Goal directed voluntary movements
- Manipulation of objects
- Integration of different senses that allows for understanding a single concept

Problems

- Inability to attend to more than one object at a time
- Inability to name an object (Anomia)
- Inability to locate the words for writing (Agraphia)
- Problems with reading (Alexia)
- Difficulty with drawing objects
- Difficulty in distinguishing left from right
- Difficulty with doing mathematics (Dyscalculia)
- Lack of awareness of certain body parts and/or surrounding space (Apraxia) that leads to difficulties in self-care.
- Difficulties with eye and hand coordination
- Inability to focus visual attention

4.2.4 Temporal Lobe

Temporal lobes are placed below the parietal and frontal lobes on either side of brain. It mainly deals with sensory input (smell, taste, voice, etc). All the auditory information from ear is received and interpreted in this part of brain, therefore called as house for auditory information. It also interprets all the collected information by our nose. Wernickes area is the main functioning area of our brain. This area is specialized in providing a person with an ability to recognize speech and interpret meanings of the words spoken. Lesions in left temporal causes disturbance with the recognition of words. whereas lesions in right hemisphere results in loss of inhibition of talking. These are highly linked with memory skills of a person. They hold a major part in long term memory eg. remembering dates, autobiographical information or places. anterograde amnesia can be one of the result if this lobe suffering from any kind of damage i.e. inability to create new memories. Recall of non-verbal material and impaired verbal material are the major problems which can be faced by damage caused to this lobe. Some of the main functions and problems of this lobe are as under:

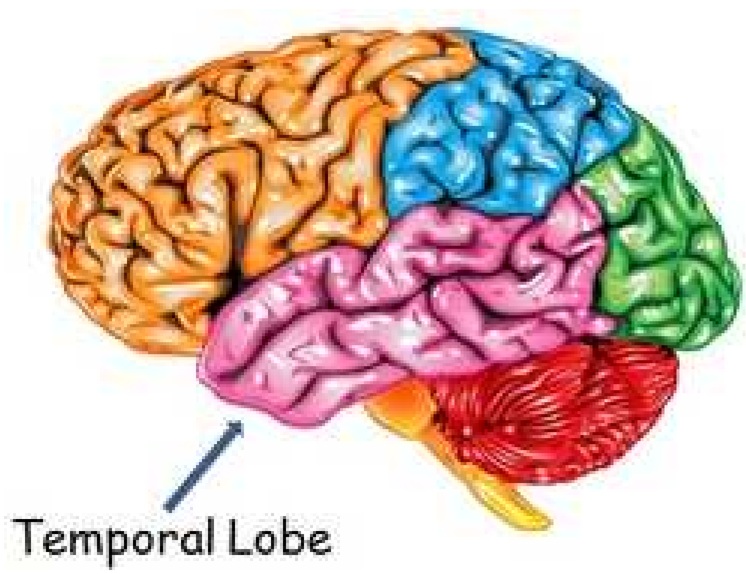

Figure 4.3: Temporal Lobe

Functions

- Hearing ability
- Memory acquisition
- Some visual perceptions
- Categorization of objects

Problems

- Difficulty in recognizing faces (Prosopagnosia)
- Difficulty in understanding spoken words (Wernicke's Aphasia)
- Disturbance with selective attention to what we see and hear
- Difficulty with identification of and verbalization about objects
- Short-term memory loss. Interference with long-term memory increased or decreased interest in sexual behavior
- Inability to categorize objects (Categorization)
- Right lobe damage can cause persistent talking
- Increased aggressive behavior

4.2.5 Occipital Lobe

The occipital lobe is located behind the temporal and parietal lobes, in the rear portion of the skull. Primary visual cortex is the most important part of this lobe. All the inputs from retina are received by this part. It is used to interpret the meaning of all visual messages received from retina. Colour and other visual aspects are also interpreted by this specialized area. Visual Images of Language are also received in the Visual receiving area. And these signals are also interpreted. Visual association area is the part where it is interpreted. This lobe is critically important in reading. This lobe is not particularly vulnerable to physical injury. Homonomous loss of vision can be caused by damage to one side of the occipital lobe which is causes with exactly the same "field cut" in both eyes. Disorders of the occipital lobe can cause visual hallucinations and illusions. Visual illusions (distorted perceptions) can take the form of objects appearing larger or smaller than they actually are, objects lacking color or objects having abnormal coloring.

Function

- Responsible for processing visual information from the eyes

Problem

- Difficulty with locating objects in environment
- Difficulty with identifying colors (Color Agnosia)
- Production of hallucinations Visual illusions - inaccurately seeing objects

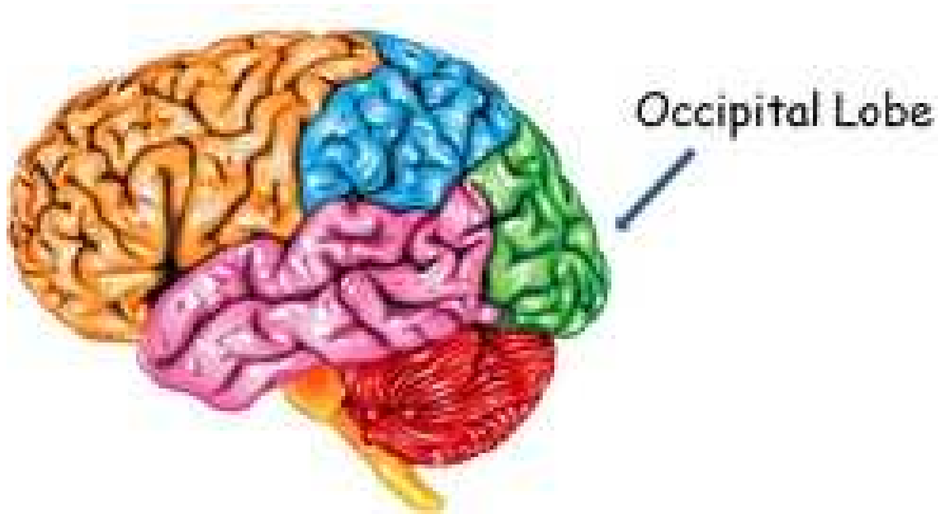

Figure 4.4: Occipital Lobe

- Word blindness - inability to recognize words
- Difficulty in recognizing drawn objects
- Inability to recognize the movement of an object (Movement Agnosia)
- Difficulties with reading and writing

4.3 Dataset Evaluation

The main dataset contains 14 parameters which predict the eye state. But, feature importance need to be accessed so that only important features are included. Features selection is done using GINI Index. The vital features of data set are listed in Table 4.5 :

Table 4.5: Dataset Parameters.

SN	Feature	Detail
1	O1	Occipital (left Hemisphere)
2	F7	Frontal (left Hemisphere)
3	P7	Pareital (left Hemisphere)
4	T8	Temporal (Right Hemisphere)
5	F8	Frontal (Right Hemisphere)
6	FC5	Between F and C(left Hemisphere)
7	FC6	Between F and C (Right Hemisphere)

4.3.1 Model Evaluation factors

Accuracy

Accuracy is a statistical measure to estimate closeness between predicted values and standard values. It is evaluated between the actual values and the predicted values by the respective models. Accuracy will be determined.

$$Accuracy = \frac{(\sum)}{n} * 100$$

$$f_n = \begin{cases} 1, & \text{if } actualvalue = predictedvalue \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

ROC Value

It is graphical representation of accuracy on a X-Y plot . In this curve, Area Under the Curve (AUC) value determines where the model is excellent, good or worthless for a dataset. ROC curve illustrates:-

- The tradeoff between sensitivity and specificity.
- Measure of accuracy by the AUC value.

Determination of efficiency of model is done in the following way:

$$0.90 =< AUC < 1 = excellent$$
$$0.80 =< AUC < .90 = good$$
$$0.70 =< AUC < .80 = fair$$
$$0.60 =< AUC < .70 = poor$$
$$0.50 =< AUC < .60 = fail$$

K-value

Cohen Kappa value is a statistical measure agreement for quantitative items. It is used to place the items into their respective group as according to the agreement. It is more robust than normal ratings.

Confusion Matrix

The confusion matrix of size n x n associated with a classifier shows the predicted and actual classification, where n is the number of different classes(here n=2). The accuracy and error can be determined from confusion matrix by the following formula[14] (where p is the number of correct negative predictions; q is the number of incorrect positive predictions; r is the number of incorrect negative predictions; s is the number of correct positive predictions). The following table shows the general concept of confusion matrix:

Table 4.6: Confusion Matrix

	PREDICTED NEGATIVE	**PREDICTED POSITIVE**
ACTUAL NEGATIVE	p	q
ACTUAL POSITIVE	r	s

The following are the equations used for calculating accuracy and error:

$$Accuracy = \frac{(p+s)}{(p+q+r+s)} \tag{4.2}$$

$$Error = \frac{(q+r)}{(p+q+r+s)} \tag{4.3}$$

Sensitivity and Specificity

The sensitivity is parameter defined as the probability of true results out of the number of samples which were actually true. The specificity is defined as the percent of negative positives that are actually negative.

Chapter 5

Methodology and Models

5.1 Approach

The approach of prediction of eye state is described in Fig 5.1 consists of 4 phases. The first phase is data set collection. The EEG data set consisting of nearly 15000 records of 14 EEG electrodes is collected from UCI archive. In the Second phase, we are eliminating the missing and duplicate entries in the data set followed by feature selection. Before performing feature Selection on the dataset ,we have implemented four machine learning models on our raw data set, so that we can cross check its accuracy and AUC value. If the AUC is close .99, then feature extraction is required to be performed. Features are extracted to enhance the quality of dataset and thereby requiring less time for execution. It is done using GINI Index , so that accuracy, AUC value may be greater than 0.9 but not be close to 0.99. 7 attributes are selected from dataset by feature selection as shown in table in chapter 4. In the third phase, this featured data set is trained and tested with 13 machine learning predictive models (listed in Table 5.1. Results obtained from this phase are analyzed and evaluated in the fourth phase on certain performance parameters (i.e. Sensitivity, Confusion matrix, Kappa value, Specificity, Accuracy and Receiver operating characteristics (ROC) curve) and then, k- fold validation is applied on best models (shown in Fig. 6.1), followed by ensembling of these models.

5.2 Models

The following table depicts the various machine learning models used in this work:

5.2.1 Random forest

Random Forest is the ensemble technique to find the nearest neighbour predictor. It uses divide-conquer technique to improve efficiency. Several small weak learners are used to ensemble to form one strong learner,as shown in figure 5.2. Being a strong learner ,it can

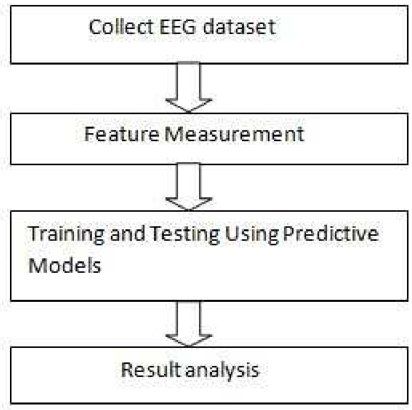

Figure 5.1: Methodology Used.

Table 5.1: Model Details

SN	Method	Model	Package
1	Wsrf	Weighted subspace random forest	Wsrf
2	Brnn	Bayesian Regularization for Feed-Forward Neural Networks	Brnn
3	Ada	Boosted classification tree	Ada Boost
4	rpart	Decision tree	Rpart
5	Nnet	Neural network	Nnet
6	Multinom	Linear model	Car,Nnet
7	Lm	Fitting linear models	Stats
8	Gbm	Stochastic Gradient Boosting	Gbm
9	Random Forest	Random forest	Random Forest
10	Ksvm	SVM	Kernlab, Hmeasure
11	Earth	Earth	Earth
12	Fda	Flexible Discriminant Analysis	MDA
13	ctree	Party	Party

efficiently run on large databases. It is learnt as error rate of random forest depends on two main factors : Correlation between to trees and strength of each individual tree[15]. More the correlation , more will be the error rate between trees and more the individual strength lesser will be the error rate. Some key Features for our Random Forest are listed below:

- It is unexcelled in accuracy among current algorithms.
- It is very effective for large databases.
- It can handle thousands of input variables without variable deletion.

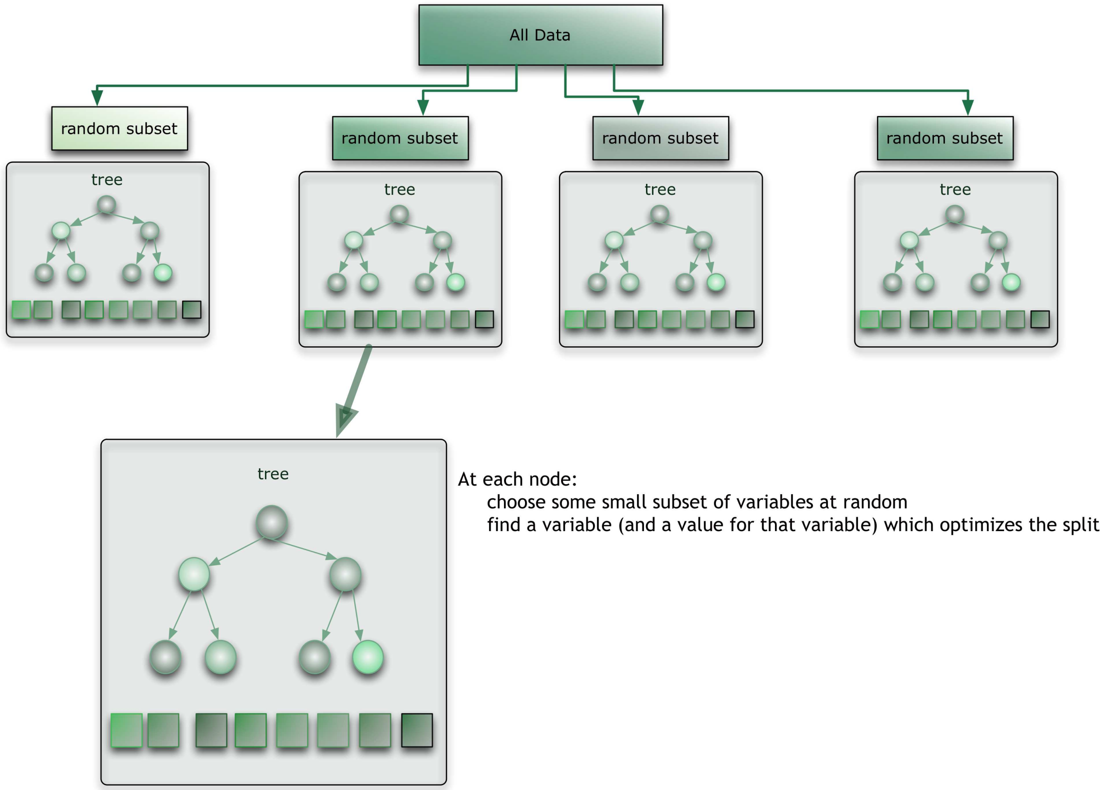

Figure 5.2: Random forest

- Accuracy maintenance and missing value estimation can be efficiently (with large number of data values missing)done.
- New databases can reuse the already saved forests formed from older data sets.
- Prototypes are computed that give information about the relation between the variables and the classification.
- Interactions of detecting variable can be examined using experimental method offered by random forest.

5.2.2 Neural Network

Neural Network are processing devices (either algorithms or actual hardware) that are loosely modeled after the neuronal structure of the mamalian cerebral cortex but on much smaller scales[16]. A large Neural Network have hundreds or thousands of processor units. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. Neural networks are also good with data sets that are noisy or where some inputs have missing variables. Each input point is a high-dimensional vector. The neural network is organized in a series of layers ,as shown in figure 5.3, where the input vector enters at the left side of the network, which is then projected to a hidden layer. Each unit in the hidden layer is a weighted sum of the values in the first layer. This layer then projects to an output layer, which is where the desired answer appears. The network is trained with the input and the desired output, which in our case are eeg parameters and eye state is our final outcome. Main disadvantage of neural network is that, it can take time in training than other machine learning models. Some of the Key features of neural network are listed below:

- Neural networks are also good with data sets that are noisy or where some inputs have missing values.
- It can easily train large input dataset and output values,which are not provided with any mapping function.
- Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
- Self-Organisation: An ANN can create its own organisation or representation of the information it receives during learning time.
- Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
- Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.

5.2.3 GBM

Gradient boosting machines are a family of powerful machine-learning techniques that have shown considerable success in a wide range of practical applications for both regression and classification datasets. It can be highly customized according to the needs of the customer. This model provides the user with an ensembled prediction model from the weak models. It builds model in a stage-wise manner just as function boosting performs and it generalizes them by allowing optimization of an arbitrary differentiable loss function. The idea of gradient boosting originated in the observation by Leo Breiman [17] and freund[18] that boosting can be interpreted as an optimization algorithm on a suitable cost function. GBM the learning procedure consecutively fits new models to

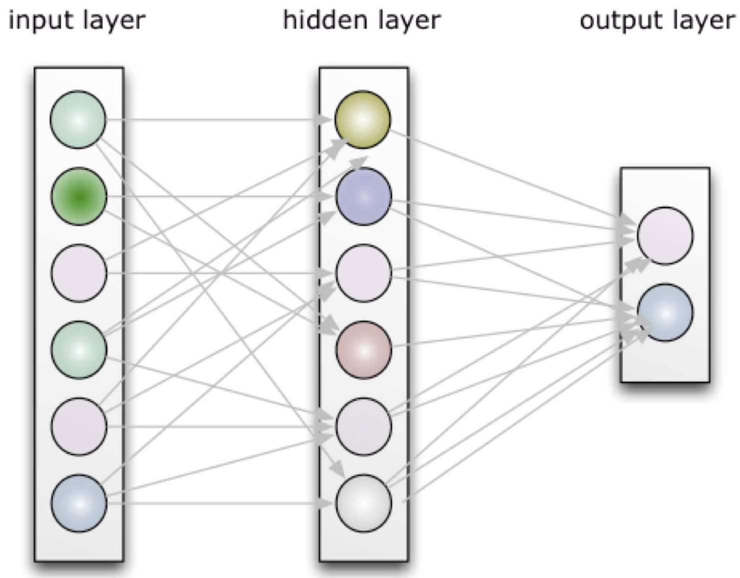

Figure 5.3: Neural Network

provide a more accurate estimate of the response variable. The principle idea behind this algorithm is to construct the new base-learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble. some of the key features of gbm are:

- It follows stepwise procedure to build model.
- They can be highly flexible for any data driven task.
- It builds new base learners, so as to increase flexibility.
- Simple to implement.
- These can be easily customized according to varied requirements of the users.

5.2.4 Decision Tree

They are known as glass-box models, because after finding patterns in the given dataset you can easily get to know what decisions will be made by this model. These are easily understandable by the ones having little experience in machine learning[19]. Decision Tree is a method for graphical representation to represent choices and results of predictive model. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. Some of the examples of decision tree are: prediction where the email is spam or not,predicting whether patient has cancer or not,crediting loan will be good or not, etc.

Initialize $\hat{f}(\mathbf{x})$ to be a constant, $\hat{f}(\mathbf{x}) = \arg\min_\rho \sum_{i=1}^{N} \Psi(y_i, \rho)$.
For t in $1, \ldots, T$ do

1. Compute the negative gradient as the working response

$$z_i = -\frac{\partial}{\partial f(\mathbf{x}_i)} \Psi(y_i, f(\mathbf{x}_i)) \Big|_{f(\mathbf{x}_i) = \hat{f}(\mathbf{x}_i)} \tag{1}$$

2. Fit a regression model, $g(\mathbf{x})$, predicting z_i from the covariates $\mathbf{x}_i$.
3. Choose a gradient descent step size as

$$\rho = \arg\min_\rho \sum_{i=1}^{N} \Psi(y_i, \hat{f}(\mathbf{x}_i) + \rho g(\mathbf{x}_i)) \tag{2}$$

4. Update the estimate of $f(\mathbf{x})$ as

$$\hat{f}(\mathbf{x}) \leftarrow \hat{f}(\mathbf{x}) + \rho g(\mathbf{x}) \tag{3}$$

Figure 5.4: Algorithm of gbm

Syntax: ctree (formula,data)

Rpart() package is used to build trees, which uses all the attributes for construction while considering dependent attributes as separate from the rest individual attributes. A model is created is created from trained (observed) data, which is then a set os validation data is used to verify/improve the results. nodes of tree formed represents the choices or results and the edges shows the rules or conditions to be fulfilled to get results.

5.2.5 Brnn

Brnn stands for Bayesian-regularized neural network . These are based on neural network with a major difference that unlike conventional approach of optimal distributing of set of weights for error minimization in neural networks, brnn uses the probabilistic distribution of networks resulting in the predictions of the network are also probability distributions. This model reduces the problem of overfitting and overtraining[20][21]. This model fits a two layer neural network. Nguyen and Widrow algorithm are used for weight assignment and for achieving optimization Gauss- newton algorithm is used. BRNN can be used in modeling highly nonlinear systems with time-series characteristics. It can be trained simultaneously in positive and negative time direction.

5.2.6 Linear Model

R model makes it very simple to fit a "Linear Model" to your statistics. R makes construction of linear models very easy. Things like "Dummy Variables", Categorical Features, "Interactions", and Multiple Regression all these come naturally[22]. In R model the center part for linear regression is the "LM" function. LM function normally comes with base R. That's why we don't have to install anything like packages and no need to import anything. We can simply fit the model to the given information by creating a formula and then passing it to the lm function.

5.2.7 WSRF Model

The "Weighted Subspace Random Forest" model was proposed in the "International Journal of Data Warehousing and Mining", and proposed by "Baoxun Xu", "Joshua Zhexue Huang", "Graham Williams", "Qiang Wang", and "Yunming Ye" [23] . The model can categorize very high-dimensional information with random forests built using small subspaces[24]. A novel variable weighting method is used for variable subspace selection in place of the traditional random variable sampling. This new approach is particularly useful in building models from high-dimensional information.

5.2.8 Party Model

Basically "Conditional Inference Trees" (C Tree) evaluate a retrogression relationship by binary recursive partitioning in a conditional inference framework. The method works in the following steps: First of all it will evaluate the global null hypothesis of independence between any of the input variables and the response. After that it will stop if this hypothesis cannot be rejected or otherwise it will select the input variable with strongest association to the response[25]. This association is measured by a p-value corresponding to a test for the partial null hypothesis of a single input variable and the response. In the second step it will implement a binary split in the selected input variable. In the third step it will recursively repeat steps first and second.

5.2.9 Earth Model

The earth R package constructs "Regression Models" using the methods in Friedman's papers "Multivariate Adaptive Regression Splines" [26] and "Fast MARS" [27] [28]. The package can be downloaded from the given link http://cran.r-project.org/web/packages/earth/index.html. The term ”MARS” is trademarked and thus not used in the name of the package. A backronym for ”Earth” is ”Enhanced Adaptive Regression Through Hinges”. The following aspects of MARS are mentioned in Friedman’s papers but not implemented in earth:

- Piecewise cubic models
- Model slicing
- Handling missing values.
- Automatic grouping of categorical predictors into subsets.
- The h parameter of Fast MARS.

5.2.10 Support Vector Machine

Support Vector Machines were developed by ”Cortes” and ”Vapnik” in the year 1995 [29] for binary classification. Their method may be roughly sketched as follows:

Class Separation Basically they are looking for the Optimal Separating Hyper plane between the two classes by maximizing the margin between the classes.

Overlapping classes Data points on the wrong side of the discriminated margin are weighted down to reduce their influence.

Nonlinearity When we cannot find a linear separator, data points are projected into a higher-dimensional space where the data points effectively become linearly separable.

The R interface to ”Libsvm” in package e1071, support vector machine, was designed to be as intuitive as possible. Models are fitted and new data are predicted as usual, and both the vector/smatrix and the formula interface are implemented [30]. As expected for Rs statistical functions, the engine tries to be smart about the mode to be chosen, using the dependent variables type (y): if y is a factor, the engine switches to classification model otherwise, it behaves as a regression machine; if y is omitted, the engine assumes a novelty detection task.

5.2.11 Fitting Linear Model

LM function is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance. Models for lm are specified symbolically. A typical model has the form response terms where response is the numeric response vector and terms are a series of terms which specifies a linear predictor for response. A terms specification of the form first and second indicates all the terms in first together with all the terms in second with duplicates removed. A specification of the form first: second indicates the set of terms obtained by taking the interactions of all terms in first with all terms in second. The specification first multiplied by second indicates the cross of first and second.

Chapter 6

Result and Discussion

Machine learning is a data analysis method. Each models follows some specific algorithm that learns from dataset provided and allows computers to find hidden observation without being explicitly programmed where to look. Various Machine learning models are used to explore different types of dataset (i.e. classification or regression). These models are like black boxes for the user to solve the problem. Each model follows their defined rules and algorithm on different dataset and their performance will vary according to the dataset selected. Classification dataset are those which have their predictive parameter in the form of real value only i.e. 1, 2, 3 etc; whereas in regression dataset prediction parameter ranges vastly over the number system. As our dataset is of binary classification type ,i.e. prediction parameter have only two values (0 and 1). we have used 13 classification dataset type models or dual use models (listed in table 5.1) to predict the eye state results on the basis of Sensitivity, Confusion matrix, Kappa value, Specificity, Accuracy and Receiver Operating Characteristics (ROC) curve. The results of all the models are shown in Table 6.1. WSRF model gives us maximum accuracy value for our dataset, followed by random forest model and then model Stochastic Gradient Boosting(GBM). These three machine learning models are used for ensembling, explained in section 6.1 .

6.1 ENSEMBLING OF MODELS

Ensembling of models means running two or more data analytic models but getting the results in a single parameter, so that we can have better prediction analysis. By ensembling different models data analysts can remove some drawbacks which are caused when they are individually run on the dataset. It is a supervised learning technique to provide a high degree of robustness. It performs best when correlation value is low i.e. models that are to be ensembled have very low linking value. It is not a good idea to ensemble models with high correlation value. Ensembling models give better prediction values and more stable model with decreased number of drawbacks. Hence providing with a better decision making parameter basis. One best example of ensembling is Random forest models, these are combination of various layers of decision trees. All the 13 predictive models used on our binary type classification dataset. It can also be defined as ensembling is performed to obtain a new model by combining 2 or more pre-defined predictive models, so as to obtain a new accuracy whose range is more defined and precise than the combined accuracy range of predictive models selected.

In this thesis, top three models selected for ensembling are WSRF, Random Forest and GBM from the comparison of models, as performed in section **??**. These selected models are ensembled by calculating accuracy from single prediction value, which is obtained by merging prediction values of our predictive models. Accuracy obtained from ensembling is 94.28.

6.2 K- FOLD VALIDATION

K- Fold validation is used to evaluate predictive models. It is also called rotation estimator. K-Fold validation is done to compare accuracy values of the prediction model. In K-Fold validation, main sample is divided into k equal sized parts, where one sub-part is taken for testing the models and the all other parts are used for training data phase. This process is similarly applied to all the remaining the k-1 subparts as for testing the model individually and their respective k-1 subparts for training the data, this is also said as cross validation phase. The main advantage of this validation is that is that every subpart is separately selected for training and tested phases. Every data set get checked in testing phase exactly once.It is mainly used to guard Type 3 errors i.e. testing hypotheses suggested by the data. It can also be used in variable selection. we can graphically represent the cross validation using different graphs. another advantage of cross validation is that we can choose how large a dataset has to be. Figure 6.1 shows the cross validation done on the ensembled models in the previous section. In this work, we have done the cross validation on the ensembled model, obtained by ensembling WSRF, Random Forest and gbm. Scatter plot is used to graphically represent the results from the cross validation (as shown in Fig 6.1).

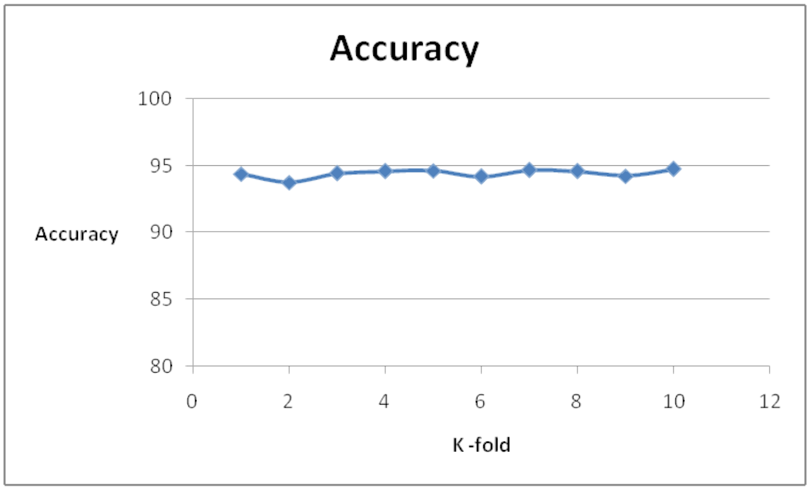

Figure 6.1: K-Fold Cross Validation.

Table 6.1: Comparative Performance.

SN	MODEL	Accuracy	Sensitivity	AUC Value	Specificity	Error
1	adaBoost	80.79	0.73	0.885	0.871	0.192
2	decision tree	71.65	0.731	0.768	0.705	0.284
3	neural network	63.72	0.469	0.675	0.775	0.363
4	linear model	64	0.483	0.676	0.768	0.36
5	Random Forest	90.82	0.862	0.969	0.946	0.092
6	Svm	80.457	0.741	0.883	0.857	0.195
7	Wsrf	95.92	1	0.999	0	0.55
8	Earth	72.073	0.641	0.783	0.786	0.279
9	gbm	88.45	0.852	0.953	0.911	0.115
10	mda	72.89	1	0.72	0	0.55
11	party	81.28	0.756	0.881	0.859	0.187
12	brnn	67.11	0.501	0.731	0.811	0.329
13	Fitted Linear Model	63.44	0.481	0.668	0.759	0.366

Chapter 7

Conclusion

In this study ,we have taken an EEG dataset to predict the state of eye by exploring 13 machine learning models on our dataset. The dataset consists of 14 eeg parameters and one prediction parameter. By analyzing the results we got WSRF, Random Forest and GBM as our best prediction models, on the basis of Sensitivity, Confusion matrix, Kappa value, Specificity, Accuracy and Receiver operating characteristics (ROC) curve. Top 3 models are then ensembled into a new model giving ensembled accuracy as 94.28, which results into more precise and highly consistent accuracy value. Cross validation is also performed on prediction models so as to obtain a mean response value. Future scope of this work is that it can be extended by including more eye states.

Publication

7.1 Publication

1. Dipali Singla, Prashant singh Rana, ”Eye state Prediction using Ensembled Machine Learning Models” to 2016 International Conference on Advances in Computing, Communications and Informatics.(Status - communicated).

7.2 Video

https://www.youtube.com/watch?v=FT5r9pr6igs

References

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

[2] Erkang Cheng, Bin Kong, Rongxiang Hu, and Fei Zheng. Eye state detection in facial image based on linear prediction error of wavelet coefficients. In *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, pages 1388–1392. IEEE, 2009.

[3] Mervyn VM Yeo, Xiaoping Li, Kaiquan Shen, and Einar PV Wilder-Smith. Can svm be used for automatic eeg detection of drowsiness during car driving? *Safety Science*, 47(1):115–124, 2009.

[4] Osamu Fukuda, Toshio Tsuji, and Makoto Kaneko. Pattern classification of eeg signals using a log-linearized gaussian mixture neural network. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 5, pages 2479–2484. IEEE, 1995.

[5] Mridu Sahu, NK Nagwani, Shrish Verma, and Saransh Shirke. An incremental feature reordering (ifr) algorithm to classify eye state identification using eeg. In *Information Systems Design and Intelligent Applications*, pages 803–811. Springer, 2015.

[6] S Natarajan, KNB Murthy, et al. A data mining approach to the diagnosis of tuberculosis by cascading clustering and classification. *arXiv preprint arXiv:1108.1045*, 2011.

[7] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240. ACM, 2006.

[8] Xue-wen Chen and Michael Wasikowski. Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 124–132. ACM, 2008.

[9] Mehdi Naseriparsa, Amir-Masoud Bidgoli, and Touraj Varaee. A hybrid feature selection method to improve performance of a group of classification algorithms. *arXiv preprint arXiv:1403.2372*, 2014.

[10] Taghi M Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. Comparing boosting and bagging techniques with noisy and imbalanced data. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(3):552–568,

2011.

[11] Tian-Yu Liu. Easyensemble and feature selection for imbalance data sets. In *Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS'09. International Joint Conference on*, pages 517–520. IEEE, 2009.

[12] Kehan Gao, Taghi M Khoshgoftaar, and Randall Wald. Combining feature selection and ensemble learning for software quality estimation. In *The Twenty-Seventh International Flairs Conference*, 2014.

[13] Huan Liu, Hiroshi Motoda, Rudy Setiono, and Zheng Zhao. Feature selection: An ever evolving frontier in data mining. *FSDM*, 10:4–13, 2010.

[14] Sofia Visa, Brian Ramsay, Anca L Ralescu, and Esther Van Der Knaap. Confusion matrix-based feature selection. In *MAICS*, pages 120–127, 2011.

[15] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[16] Michael T Wishart and Ronald G Harley. Identification and control of induction machines using artificial neural networks. *Industry Applications, IEEE Transactions on*, 31(3):612–619, 1995.

[17] Wenxin Jiang. Some theoretical aspects of boosting in the presence of noisy data. In *Proceedings of the Eighteenth International Conference on Machine Learning*. Citeseer, 2001.

[18] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[19] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[20] Jae-Young Kim, Dong-Chul Park, and Dong-Min Woo. Electric load prediction using a bilinear recurrent neural network. In *Computer Modelling and Simulation (UKSim), 2010 12th International Conference on*, pages 404–407. IEEE, 2010.

[21] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, 45(11):2673–2681, 1997.

[22] Christian F Beckmann and Stephen M Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 23(2):137–152, 2004.

[23] UK10 Hayes, Shyue-Liang Wang, Chung-Yi Chen, I-Hsien Ting, Tzung-Pei Hong, Damla Oguz, and Turkey10 Izmir. International journal of data warehousing and mining. 2013.

[24] G Williams Q Wang YB Xu, JZ Huang and Y Ye. Weighted subspace random forest for classification, October 2015.

[25] Achim Zeileis, Torsten Hothorn, and Kurt Hornik. Model-based recursive partition-

ing. *Journal of Computational and Graphical Statistics*, 17(2):492–514, 2008.

[26] Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.

[27] Stanford University. Laboratory for Computational Statistics and Jerome H Friedman. *Fast MARS.* 1993.

[28] Stephen Milborrow. Notes on the earth package, 2014.

[29] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[30] Yonas B Dibike, Slavco Velickov, Dimitri Solomatine, and Michael B Abbott. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3):208–216, 2001.