Schizophrenia detection using ELM

Major project report

By

SOMNATH SAHA(36), ABHINAV KESARWANI(02)

Department of Computer Science
UNIVERSITY OF DELHI

Research Project being submitted in the partial fulfillment of the requirement for the award of the degree of M.SC. (CS), is a record of original work carried out by the undersigned in the Department of Computer Science, University of Delhi.

MAY 2018

CERTIFICATE

This is to certify that the project report titled "Schizophrenia Detection using ELM", submitted by Somnath Saha and Abhinav Kesarwani to the Department of Computer Science, University of Delhi, is a bonafide record of the work done by them under the supervision of Dr. Naveen Kumar and Mr. Ankit Rajpal, for the Major Project of M.Sc. Computer Science (2016-2018).

Dedication and acknowledgements

We are thankful to our supervisor, Dr. Naveen Kumar, Department of Computer Science, Delhi University, for his valuable contributions and generous suggestions towards the completion of this project. He always strive to provide us with best of his capabilities. We feel honored and privileged to work under him. We are thankful to him for spending his time with us and guiding us at each and every step and his immense pool of knowledge and expertise of the topic. I am also immensely grateful to Mr. Indranath Chatterjee for his constant support and the lab staff of the Department of Computer Science, University of Delhi for providing us with lab facilities for carrying out the experiments. We truly believe that this project would not have been possible without the support of all of them.

Author's declaration

We declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Master's Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Somnath Saha(36)
Abhinav Kesarwani(02)

TABLE OF CONTENTS

List of Tables

Table **Page**

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

In the introduction part of the report we have discussed about the Schizophrenia disease, Computer aided diagnosis, Neural network, introduction to ELM, problem statement and objective of study.

1.1 What is Schizophrenia?

Schizophrenia is a psychiatric disorder in which a person shows some positive, some negative and cognitive symptoms. Here positive doesn't means that a person will show some positive sign instead it means a person will show signs some symptoms which are normally not present in human body. Positive symptoms are psychotic in nature like delusions, hallucination, disorganized speech, disorganized behavior, catatonic behavior. Like for example a TV reporter said "There are chances of earthquake", so a person might think he is directly referring to him. Hallucinations can be a person thinking there might be another person standing beside him though he might not be physically present at the moment.Schizophrenia has 3 cyclic phases:

1. Prodromal phase in which a person is withdrawn from the social

environment.

2. Active phase in which a person shows positive symptoms.
3. In Residual Phase person shows cognitive symptoms like memory, learning and understanding gets affected.

Currently, around 1.1% of world's total population is suffering from Schizophrenia which is around 51 people around the world. [Source]

1.2 What is Computer-aided diagnosis (CAD)?

Computer-aided diagnosis (CAD) has become one of the major research subjects in medical imaging and diagnostic radiology. Computer-aided diagnosis tool is used for automated classifications and also can identify most distinct voxels or features. Which when backtracked to brain can identify affected brain regions. Functional magnetic resonance imaging(fMRI) helps to design an automated tool for diagnosis of schizophrenia. It is a neuroimaging technique that captures brain activity in small units of the brain volume called voxels, by measuring the change in blood-oxygen-level dependent (BOLD) signals over time. This change in blood oxygen level plays an important role in detection of schizophrenia. The difference in the magnetic properties causes small differences in the magnetic resonance (MR) signal of blood depending on the degree of oxygenation. This fMRI is done using different types of task related activities. Many models based on machine learning are proposed to investigate fMRI data to detect schizophrenia. High dimensionality is a big problem for machine learning algorithms to fMRI data. The fMRI data is a time series data. So it consists of many 3D brain images over time. A 3D fMRI image is thought of a sequence of 2D images(slices) across the whole brain. In every 2D image are divided into small units called voxels. But due to huge

number of voxels and also for redundancy and irrelevancy of features, the classification algorithms do not work accurately. For this reason, feature selection is needed.

1.3 Supervised Learning

Supervised learning is a type of algorithm of machine learning which uses labeled data to make predictions. The data is divided into two parts : training data and test data. The training data includes input data and response values. Supervised learning techniques build models using training data set so that it can predict the response values of the test dataset. After this the test set data is used to validate the model. Larger datasets results in better models which have a higher predictive power and can very well predicts values for the new dataset. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers. The algorithm iterates recursively on the training set and make predictions which is then corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

1.4 The Problem Statement

Till 2015, 17000 people have died due to schizophrenia. Currently there is no physical or lab test available to detect schizophrenia. Mostly it is diagnosed over a period of 6 months (with 1 month of active symptoms) by a psychiatrist based on clinical symptoms. If we can automate this Schizophrenia detection, then this will help to fight against the problem. One of the fast approach is to use Machine Learning technique in order to make a system learn and then work by itself. Our aim of this project is to classify schizophrenic patients faster than the current diagnoses available

and maximize the accuracy of the detection. We have a dataset called "Function Biomedical Informatics Research Network Data Repository" i.e. FBIRN which is obtained from the Function BIRN Data Repository. We want to use good feature selection algorithm to reduce number of features and classify a subject after training a model using classifier.

1.5 Objective of Study

Currently there is no physical or lab test available to detect schizophrenia. Mostly it is diagnosed over a period of 6 months (with 1 month of active symptoms) by a psychiatrist based on clinical tests. If we can automate this Schizophrenia detection, then this will help to fight against the problem. One of the fast approach is to use Machine Learning technique in order to make a system learn and then work by itself. Our aim of this project is to classify schizophrenic patients faster than the current diagnoses available and maximize the accuracy of the detection.

CHAPTER 2

RELATED WORKS

2.1 Related works

[Vince D. Calhoun][11] Used Functional Magnetic resonance imaging data and a multivariate analysis method, independent component analysis to determine the brain regions and discriminate between the bipolar, schizophrenic patients and healthy patients. Brain regions (Temporal lobe and default mode network) were identified in all the three types of subjects. There are total of 61 patients in which 21 are schizophrenic patients, 14 have bipolar disorder and rest 26 are healthy subjects. They used LOOCV (Leave one out cross validation) and obtained an accuracy of 98.3% in case of schizophrenic patients, 99.0% in case of bipolar disorder and 99.1% in case of healthy subjects.

Gur, Raquel E., et al (Gur, 2002)have calculated the blood oxygenated response in patients suffering from schizophrenia during facial emotion processing task to test the cause of diminished limbic activation related to emotional relevance of facial stimuli. They have used 14 (consisting of 10 men and 4 females) patients suffering from schizophrenia and 14

comparison subjects with facial expressions (happiness, anger, disgust, sadness, fear and neutral as well). The groups did not differ in performance of the task ,healthy participants showed activation in the fusiform gyrus, occipital lobe, and inferior frontal cortex relative to the resting baseline condition. While calculating the voxels uniquely activated by emotional valence discrimination task in case of healthy participants showed responsive voxels in both left amygdala and hippocampus and in case of schizophrenic patients there was a little less response in the left amygdala. The authors finally concluded that the failure to activate limbic regions during emotional valence discrimination task may explain emotion processing deficits in patients with schizophrenia.

[Cameron S. Carter][3] Author recognized anterior cingulate cortex region in schizophrenic patients while examining brain activity associated with internal monitoring of performance. A total of 33 subjects were given a task related functional magnetic resonance imaging. Stimulus degradation was used to increase the error rate. Healthy subject show error related activity in anterior cingulate cortex. While removing the error rate the patients showed slowing of reaction time.

Rubia, Katya, et al. (Rubia, 2001)have performed task on six schizophrenic patients aged between 26-57 years and seven healthy patients aged 26-58 years, not having any personal or family history of psychiatric disorder. Patients were asked to perform two motor tasks ‘stop’ and ‘go-no-go’ tasks. Both tasks were sensory motor task related to pressing button when they see something on screen. It was proposed that schizophrenic patients show decreased response in left anterior cingulate in both task. Also, they show decrease in response of left rostral dorsolateral prefrontal and increased thalamus and putamen response during stop task. Patients shows reduced left prefrontal activation and larger subcortical activation during motor task.

[Darya Chyzhyk][4] In this paper the author has taken resting state fMRI data and recognized brain regions which are affected in schizophrenic patients. The dataset used is COBRE in which there are a total of 146 patients in which 74 are healthy subjects and 72 are schizophrenic patients. They used Pearson's correlation for feature selection and then applied ensembles of Extreme learning machine to build a computer aided diagnosis system in which achieved an accuracy of 91.19%.

Manoach, Dara S., et al. (Manoach, 1999)performed the task on 12 schizophrenic and 10 normal subjects. Using fMRI , performing working memory task , activation of dorsolateral prefrontal cortex (DLPFC) was compared between the patients and healthy subjects. Subjects were made to perform a modified Sternberg Item Recognition Paradigm (SIRP) task. A high working memory load condition was compared with non working memory condition and with a low working memory load condition. They find out that during neuroimaging studies of patients suffering from schizophrenia, who performed working memory task shows abnormally diminished activity of prefrontal cortex as compared to normal subjects. It was found that schizophrenic patients performed worse than normal patients. Large activation was seen in the left DLPFC but no difference in the right DLPFC. According to measured errors , it was recorded that left DLPFC activation was inversely correlated to task performance , for schizophrenic group.

[Oguz Demirci][9] fMRI data of 155 subjects, obtained from 2 sites and using 3 different tasks are used to investigate the effectiveness between each task and tell the difference of each task on patients who are healthy and schizophrenic. Independent component analysis has been used for feature selection and then principal component analysis (PCA) has been applied. In conclusion it shows that SM task provides a better result as

compared to AOD and SIRP task.

Manoach, Dara S., et al. (Manoach, 2000) analysed region-wise brain activation in nine schizophrenic subjects and nine healthy subjects using fMRI. Subjects performed (working memory) a modified version of Sternberg Item Recognition Paradigm (SIRP), which included a cash reward for correct responses. They compared high and low working memory load conditions and also with non working memory condition. They observed region- wise brain activations in individuals as well as in groups. . It was seen that schizophrenic patients not only show weak working memory performance but also basal ganglia and thalamus were activated in only schizophrenic patients. These regions were active in group level as well when compared with normal group.

[Oguz Demirci][10] Used a task based (Auditory task ball) fMRI data for the classification of schizophrenia and healthy patients. Independent component analysis has been applied on 70 subjects to obtain the best features and then on application of novel projection pursuit technique, it gives an accuracy of about 80-90%. This model can be used in differentiation of schizophrenia and healthy patients.

Wible, C. G., et al. (Wible, 2009)They have examined verbal hallucinations in schizophrenic patients while performing a working memory task. Data from Sternberg Item Recognition Paradigm which is working memory task that the data were acquired by performing the functional magnetic resonance imaging procedures(fMRI).They have taken 74 schizophrenic subjects(20 female and 54 male).They subdivided the participants into non hallucinating and hallucinating groups. They found out that the patients with auditory hallucinations (compared to non hallucinating subjects) showed decreased activity during the probe condition in working memory including the inferior parietal regions and superior

temporal regions.

[Avram J. Holmes][2] Patients with Schizophrenia mostly show decreased cerebral blood flow (CBF) in the prefrontal cortex of the brain. However to show that it is specific to schizophrenia, the author has taken a dataset of 9 healthy subjects, 10 depressed subjects and 7 schizophrenia subjects. The behavioral patterns were seen to be consistent in case of schizophrenia while depressed subjects didn't show any signs of that. The imaging data show abnormal activity in right middle frontal gyrus in schizophrenia patients related to context processing.

Potkin, S. G., et al. (Potkin, 2009, pp. 19-31).They have performed a study on Working memory where they have examined the BOLD signal change in an atlas-based demarcation of the dorsolateral prefrontal cortex (DLPFC). In this functional magnetic resonance imaging (fMRI) study they have taken 128 schizophrenic subjects and 128 healthy matched controls. They came to result that the subjects without schizophrenia preformed slightly but significantly better than the participants with the schizophrenia and have lesser reaction times when the memory load was high. The mean BOLD signal was also found to be higher during heavy memory loads in schizophrenic subjects as compared to healthy controls.

[M.M. Machulda][8] A fMRI memory encoding task is used to distinguish between the patients with mild cognitive impairment (MCI) and Alzheimer's disease (AD). 29 subjects were taken into consideration in which 11 were healthy, 9 having MCI and 9 having AD. A passive sensory task was also perform to assess the potential intergroup difference in fMRI responsiveness. Medial temporal lobe activation was less for subjects having MCI and AD as compared to healthy one in memory encoding task. While similar activations were seen in sensory task.

[Graziella Orrù][5] Over the years univariate analysis of neuroimaging data has been performed for classification of healthy individuals and patients suffering a wide range of neurological and psychiatric disorders like Alzheimer's disease, Parkinson disease, schizophrenia, depression, bipolar disorder, Huntington's disease. Recently the attention has been turn towards alternative forms of analysis like Support Vector Machines (SVM) for diagnosis, treatment and functional neuroimaging.

[Alexandre Savio][1] Resting state fMRI data from COBRE dataset which has data of 146 patients (72 schizophrenia and 74 healthy) has been used for image biomarkers of psychiatric disorder such as schizophrenia. T-test has been applied for feature extraction and then support vector machines have been used to classify between the two groups. They used LOOCV for segregation of training and testing data and consequently obtained an average accuracy of 80%.

[Hui Shen][12] Resting state fMRI data of 52 patients have been collected for classification of schizophrenia and healthy subjects. The dataset has been obtained from Department of Psychiatry, Second Xiangya Hospital of Central South University. A total number of 6670 features have been selected using Kendall tau rank correlation method and LLE (low dimensional representation). Since there were limited number of subjects, LOOCV has been applied for training and testing and then clustering algorithms for segregation of two groups. After this they obtained high classification accuracy of 93.75% for schizophrenia and 75% for healthy subjects.

[Martha E. Shenton][6] Data from postmortem, CT scans and magnetic resonance imaging have shown that left temporal gyrus show abnormalities in case of Schizophrenia patients. A total of 30 patients have been taken into consideration in which 15 have schizophrenia and 15 are

healthy. The patients having schizophrenia show volume reduction in grey matter in left anterior hippocampus- amygdala, left parahippocampal gyrus and left superior temporal gyrus. It means volume reductions is localized in left temporal gyrus of grey matter.

CHAPTER 3

METHODS AND MATERIALS

3.1 Dataset

3.1.1 About the Dataset

All the data used for this study were obtained from the Function BIRN Data Repository. FBIRN repository contains the multi-site fMRI dataset which includes schizophrenia and healthy subjects. The data was acquired using 1.5T scanner keeping all other parameters same for the subjects. In this study, we have used BOLD fMRI data of Auditory oddball (AUD) task, where all subjects had regular hearing levels, sufficient eyesight, and were able to perform cognitive task. Healthy subjects were excluded if they had a current or past history of head injury or major medical illness. Only those subjects with schizophrenia and schizoaffective disorder were allowed who met the criteria as per the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).

In our study, we have used dataset which contains fMRI data of 30 schizophrenia patients and 30 healthy subjects (available at site 0009 and site 0010 of FBIRN repository), which were acquired with 1.5T scanner.

Table 3.1: Dataset details

	No. of Subjects	**Age group**	**Male/Female**
Healthy	30	40.4 +- 12.29	20/10
Schizophrenia	30	42.3 +- 10.81	24/6

Four runs of each subject's scan have been used for the experiments. Table 3.1 shows the demographic details of the dataset.

3.1.2 Task Details

Auditory oddball task is a common task [3, 25, 29, 36, 40] used to detect alterations in brain activation patterns that help to differentiate between schizophrenia and healthy subjects. A subject is presented with a continuous stream of sound, and he/ she must identify the sequence of discrete stimuli comprising standard tones and deviant (i.e. oddball) tones. Standard tones, i.e., 1000 Hz appear for 95% of trials. Deviant (i.e. oddball) tones (1200 Hz) that are distinct from standard tones, appear occasionally (5% of trials). The FBIRN conducted the Auditory oddball task consisting of four experimental runs, each having duration of 280 seconds. During the experiment, in each run, the subjects were asked to see a gray screen with a black fixation cross in the middle. They were asked to press button '1' each time they heard a deviant tone while focusing on the cross and listening to the tones. The task began with a fixation block of the silence of 15 seconds. Then a sequence of standard tones (duration = 100 ms) were presented. The deviant tone (duration = 100 ms) was presented every 6 to 15 seconds. A period of silence (duration = 15 seconds) ended each task run. In each experimental run, 140 brain scans were acquired with repetition time (TR) of 2 seconds.

3.1.3 Imaging parameters

According to FBIRN repository, the functional scans were T2*-weighted gradient EPI (Echo Planar Imaging) sequences. Pulse sequence parameters were closely matched based on pilot studies carried out by FBIRN research group: Orientation: anterior commissure posterior commissure line; the number of slices: 27; slice thickness: 4 mm; TR: 2 seconds ; time to echo: 40 ms for 1.5 T scanners; matrix: 64×64; field of view: 22 cm; and flip angle: 90°.

3.1.4 Data preprocessing

The raw datasets taken from FBIRN repository have been preprocessed using Statistical Parametric Mapping (SPM) toolbox version 8 (SPM8, Wellcome Trust Centre for Neuroimaging, University College London, UK).1 Raw scans were collected at voxel size of 3.4 × 3.4 × 4 mm3. These are realigned with the first scan as a reference. The slice timing correction is done to correct the possible errors by temporal variations during the acquisition of fMRI datasets. Subsequently, the fMRI scans are spatially normalized into standard Montreal Neurological Institute (MNI) space using an EPI template available in SPM8. This transforms the initial voxel's dimension to 3 × 3 × 3 mm3 and yields each volume of 53 × 63 × 46 voxels. Finally, spatial smoothing is done with a 9 × 9 × 9 mm3 full width at half maximum (FWHM) Gaussian kernel to get the smoothed volumes.

3.2 Methods

3.2.1 Ttest

The t test also called Student's T Test is a measure of comparing means which in return tells you how different they are from each other. The t test also tells you how significant the differences are that means did the

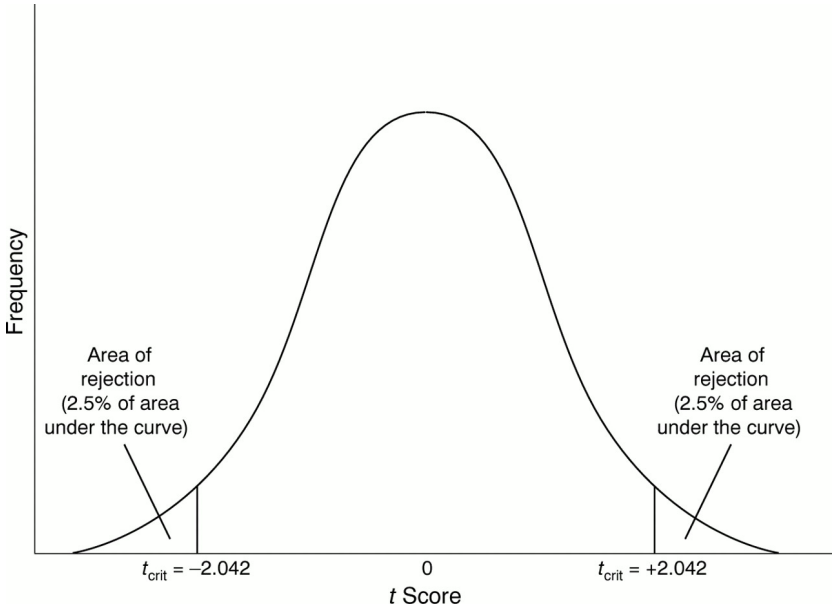

FIGURE 3.1. ttest

differences happened by chance or are they real and true. The T score T score is the ratio between the differences between the two groups and the difference within the groups. A larger t score value indicates that there is more difference between the groups. A t score of 3 signifies that the groups are three times are different from each other as they are from within.

- A large t-score tells you that the groups are different.
- A small t-score tells you that the groups are similar.

There are 3 types of t-tests

1. Independent samples- Means of two groups are compared.
2. Paired Sample t-test- Means from the same group at different times are compared.
3. One sample t-test- Given a mean we compare it with mean of single group.

When to Choose a Paired T Test / Paired Samples T Test Paired ttest is chosen when comparing same item, person or thing. Or we should choose this test if we have items which are subject to same measuring condition. Like comparing durability test of mobile phones. Though they can be from different companies still they are subject to same durability conditions. Paired sample ttest is chosen when means of two samples are being compared. For example, given two students from different universities and means of subject marks scored by both of them. The test statistics is

$$t = (X' - Y')/\sqrt{(s_x^2/n + s_y^2/m)} \tag{3.1}$$

Where X' and Y' are sample means s_x^2 and s_y^2 are the sample standard deviations n and m are the sample sizes.

3.2.2 Relief

[7] know that irrelevant features decrease the speed of the classification model because of high dimension. Also it decrease the accuracy of the model. So, we need to use a good feature selection process which will fulfill our requirements. Relief plays the role here. In relief in the initial step, we put weights to every features as 0. Then in the second step, we take a random subject(let's take it R) irrespective of it's class label. Then we find nearest subject of the same class and another nearest subject of the opposite class. We call them Near Hit and Near Miss respectively. This distances are calculated using euclidean distance. After that we update weights of all the features using this formula.

$$W(i) = W(i) - \|R(i) - NearHit(i)\|^2 + \|R(i) - NearMiss(i)\|^2 \tag{3.2}$$

This step is done as same number of times as the number of training set. After that relevance is calculated. We sort those features in descending order according to their relevancy. Then choose the best of those.

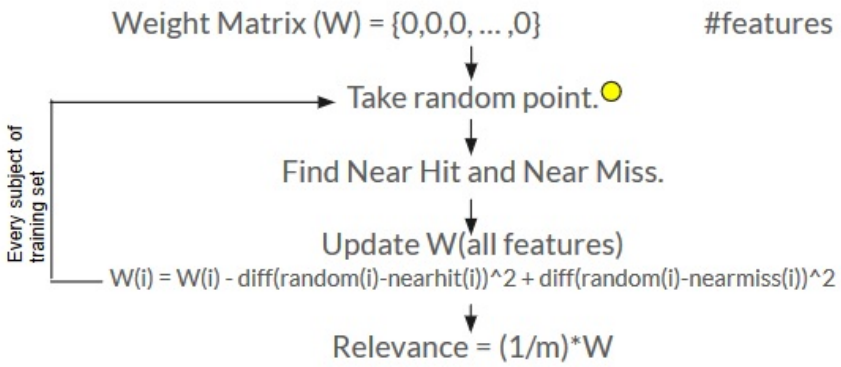

FIGURE 3.2. Relief work flow

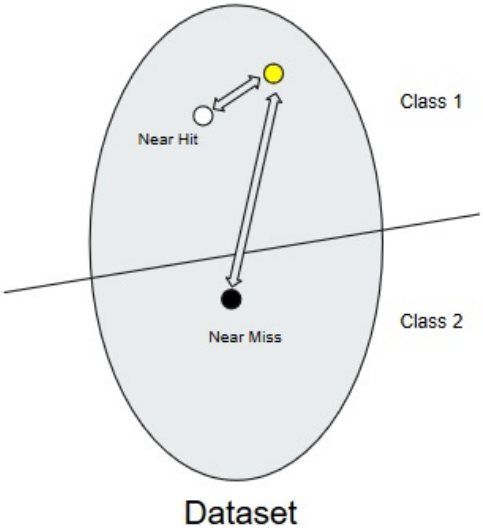

FIGURE 3.3. Relief feature selection process

3.2.3 ELM

ELM or Extreme Learning Machines is a single hidden layer feedforward neural network (SLFN) for classification, regression, clustering, sparse

approximation where hidden units not necessarily be neural units. The idea behind this method is to reduce training of the non-linear phase of the SLFN to a stochastic sampling. For N arbitrary distinct samples (x_i,t_i), where input variables are x_i and target values are t_i. The output of the generalized SLFN (Huang et al., 2015) with K number of hidden neurons and hidden unit output function h(x) is

$$f(x) = \sum \beta_i . h(x) = h(x).\beta \tag{3.3}$$

Where β is the output weights matrix of size K × m mapping the output of the hidden units to the m ≥ 1 output nodes, and the transformation h (x) is the ELM nonlinear feature mapping. (Darya Chyzhyk, 2015)

The SLFN estimation of the desired target values can be written in matrix form as:

$$H\beta = \hat{T} \tag{3.4}$$

where H, of size N × L, sometimes called the feature kernel of the ELM.

$$H = \begin{bmatrix} h(x_i) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \dots & h_L(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_N) & \dots & h_L(x_N) \end{bmatrix}$$

and $\hat{(T)}$ is the achieved approximation of the N × m target matrix.

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1\mathrm{m}} \\ \vdots & \vdots & \vdots \\ t_{\mathrm{N1}} & \dots & t_{\mathrm{Nm}} \end{bmatrix}$$

The general schema of ELM training of an SLFN has two stages: (1) construction of the feature kernel matrix H by random sampling of the parameters of the hidden unit output function, i.e. random feature mapping, and (2) estimation of the output weight matrix β by resolution of a

linear system of equations. The second stage of ELM training of an SLFN corresponds to the following minimization:

$$\min_{\beta} \|H\beta - T\|^2 \tag{3.5}$$

whose optimal solution is given by $\hat{\beta} = H^{\dagger}$T, where $H^{\dagger}$ is the Moore–Penrose generalized inverse of H. Regularization has also been introduced in order to obtain more stable and parsimonious architectures, as well as a plethora of enhanced training strategies (Huang et al., 2015).

3.2.4 Online Sequential ELM

In primitive ELM it is assumed that all the N samples of data is available at the time of training. However, in Online Sequential Extreme learning machines, the training data arrives in chunks or one by one. The primitive ELM has been modified to make OSELM.
The output weight matrix $\hat{\beta}$ is determined by,

$$\beta = H^{\dagger}\hat{T} \tag{3.6}$$

which is least square solution of $H\beta = T$. Here we consider a special case where rank(H) = N which is the number of hidden nodes. Under this condition, $H^{\dagger}$ is determined by

$$H^{\dagger} = (H^{\dagger}.H)^{-1}H^{T} \tag{3.7}$$

which is also called the left pseudo inverse of H. Now there can be a case where $H^{\dagger}.H = I_N$, hence we can make it nonsingular by making $\hat{N}(Networksize)$ smaller or increasing N in initialization phase of OS-ELM. Now after substituting the above equation in $\hat{\beta} = H^{\dagger}T$, we get

$$\hat{\beta} = (H^{\dagger}.H)^{-1}H^{T}.T \tag{3.8}$$

OSELM can be divided into two steps:

1. Boosting Phase: Given a small initial training set to boost the learning procedure. An arbitrary set of input is selected. Then calculate the hidden layer output matrix given an activation function g. Then estimate the initial output weight matrix

$$\beta^0 = K_0.H_0^T.T_0 \tag{3.9}$$

where $K_0 = (H_0^T.H_0)^{-1}$ and $T_0 = \{t_1 \cdots t_N\}^T$.

2. Sequential Learning Phase: for further observations we calculate the hidden layer output vector and then calculate the latest output weight β^{p+1} by formula

$$K_{p+1} = K_p - \frac{K_p.h_{p+1}.h_{p+1}^T.K_p}{1 + h_{k+1}^T.K_p.h_{p+1}} \tag{3.10}$$

and

$$\beta^{p+1} = \beta^p + K_{p+1}.h_{p+1}.(t_p^T - h_{p+1}^T.\beta^p) \tag{3.11}$$

3.2.5 Kernel ELM

In equation 3.7, if H has more rows than columns (N > L),which is usually the case where the number of training patterns is larger than the number of the hidden neurons, we have the following closed form solution for $\hat{\beta}$:

$$\hat{\beta} = (H^\dagger.H + \frac{I}{C})^{-1}H^T.T \tag{3.12}$$

where I is an identity matrix of dimension L.
Note that in practice, rather than explicitly inverting the L × L matrix in the above expression, we can instead solve a set of linear equations in a more efficient and numerically stable manner.
If the number of training patterns is less than the number of hidden neurons (N < L), then H will have more columns than rows, which usually gives an under-determined least squares problem. Moreover, it is less

efficient to invert a L × L matrix in this case. To handle this problem, we restrict β to be a linear combination of the rows in H: $\beta = H^T.\alpha$ ($\alpha \in R^{\mathrm{N\times m}}$). Notice that when N < L and H is of full row rank, then $H.H^T$ is invertible. Substituting $\beta = H^T.\alpha$ and multiplying both sides we get

$$\hat{\beta} = H^T(H.H^{\dagger} + \frac{I}{C})^{-1}T \tag{3.13}$$

Here, H^TH and HH^T are called ELM kernel matrix.

3.2.6 Support Vector Machine

In SVMs non-linear solutions can be efficiently found by using the "kernel trick": The data is mapped into a high-dimensional space in which the problem becomes linearly separable. The "trick" is that this is only "virtually" done by calculating kernel functions. (Link). In the simplest words, SVM algorithm, for two classes,estimate a decision boundary (a hyperplane) that separates with maximum margin a set of positive examples from a set of negative examples. Each example is an input vector X_i (i= 1,2,3...N) having features and is associated with one of the two classes Y_i = +1 or -1(Binary class). +1 belongs to the positive class and – 1 to negative class. f(x) = wx + b In this function, x refers to a training or test pattern,w is referred to as the weight vector and the value b as the bias term. The term wx refers to the dot product (inner product, scalar product), which calculates the sum of the products of vector components wixi. In case of two features, the discriminant function is, thus, simply: $f(x) = w_1x_1 + w_2x_2 + b$ decision boundary will be wx+ b =0. (link) For example, in fMRI research, our data vectors contain feature values where features are calculated upon a region(set of voxels) . If the predicted label of test data (Yi) is equal to -1 , it indicates that the data belongs to negative class, and if it is (Yi) +1 indicates condition that the data point belongs to positive class.LibSVM provides four different kernels, namely, linear kernel, polynomial kernel , radial basis function (rbf) , and sigmoid

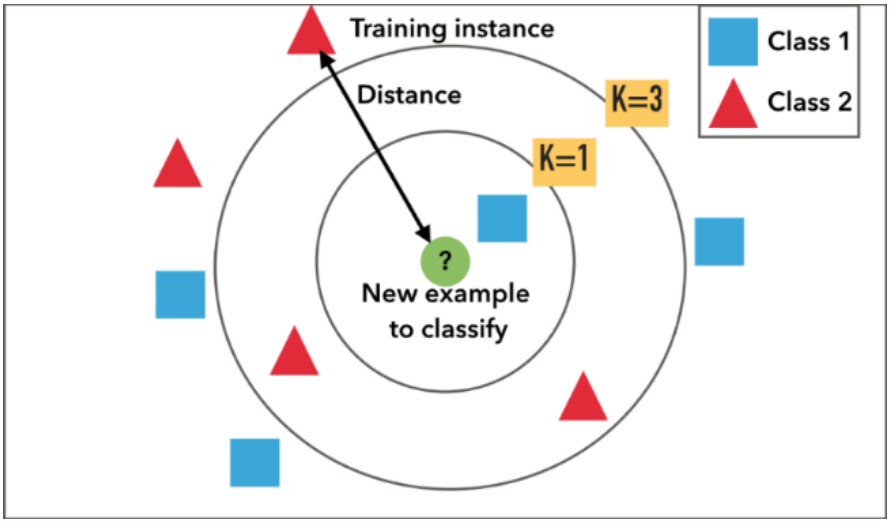

FIGURE 3.4. KNN work-flow

kernel . According to our dataset we have usedlinearl kernel and binary class. We have varied the values of c from 10 – 100 but no significant difference was found.

3.2.7 KNN

k-Nearest Neighbour (kNN) classification technique is the simplest technique conceptually. This algorithm is depends on two factors : a distance metric and a voting function , the metric we have used is the Euclidean distance, other option include minkowski, hamming etc. Voting function is calculation of majority of votes .The k-nearest neighbor classification is performed over training dataset containing values of the features. It doesn't even involve explicitly learning a classification function. Classification of a test example is done by finding the training set example that is most similar to it by some measure The test data is estimated using k = 1(by default), that is the class of the one nearest neighbor will be assigned to the test data.

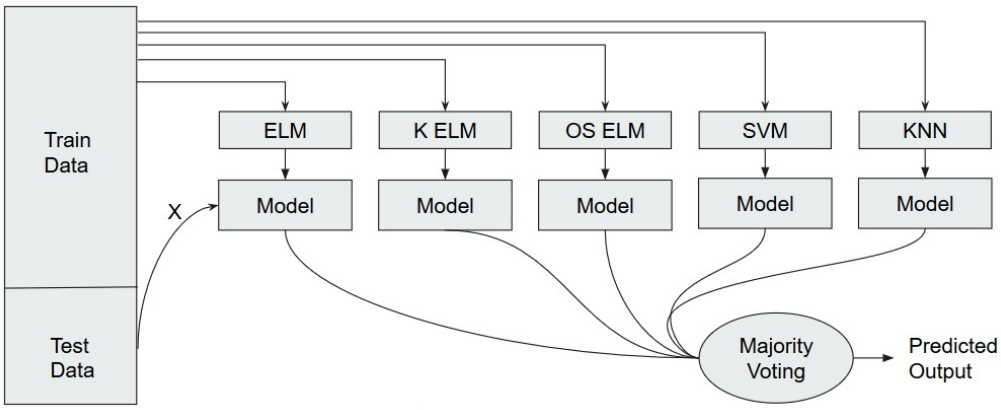

FIGURE 3.5. Ensemble of classifiers

3.2.8 Ensemble of Classifiers

In this method, the test data is sent to all classifiers and the classifiers train their models using those data. After their models are ready, test data is passed to all those classifiers independently. Those classifiers then predict the output using their models. After all the classifiers have predicted their outputs, we take the majority of outputs to determine the final output.

If you average a bunch of democratic-leaning polls and a bunch of republican-leaning polls together, you will get on average something that isn't leaning either way.

The aggregate opinion of a bunch of models is less noisy than the single opinion of one of the models. In finance this is called diversification - a mixed portfolio of many stocks will be much less variable than just one of the stocks alone. This is also why your models will be better with more data points rather than fewer.

If you have individual models that didn't overfit, and you're combining the predictions from each model in a simple way (average, weighted average, or logistic regression), then there's no room for overfitting.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results of feature selections

We have preprocessed dataset of 60*153594 matrix, where we have a record of 60 subjects in which 30 subjects are schizophrenia affected and 30 subjects are normal. Each subject has 153594 features. Out of those 153594 features, there are many redundant and/ or not useful features. For that reason, we need to reduce the number of features. We have applied statistical ttest to keep only the relevant features. For each feature, we have calculated the null hypothesis. The null hypothesis is taken as the means of 30 schizophrenia subjects and 30 normal subjects are equal. We have taken all those features whose null hypothesis got rejected. After filtering we get 2859 features. Now, all those 2859 features will not be of the same quality features. So, we have applied relief feature selection on those 2859 features. Relief gives every features a weight according to the relevancy. We have sorted those features in descending manner according to their weights.
After getting most relevant features, we have run ELM(Extreme Learning Machine) on those. We have tried taking top 10% i.e. 285 features, top

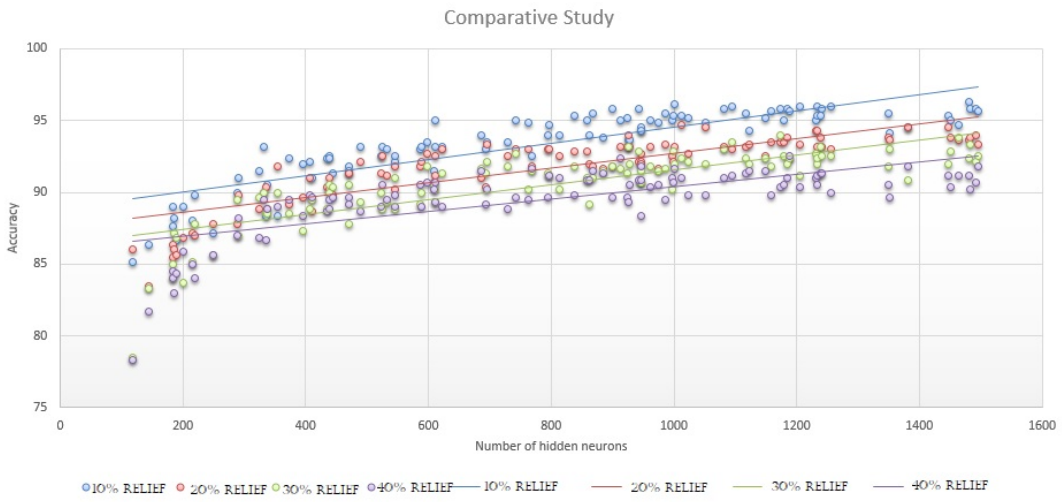

FIGURE 4.1. Comparative study between 10%, 20%, 30%, 40% feature selection

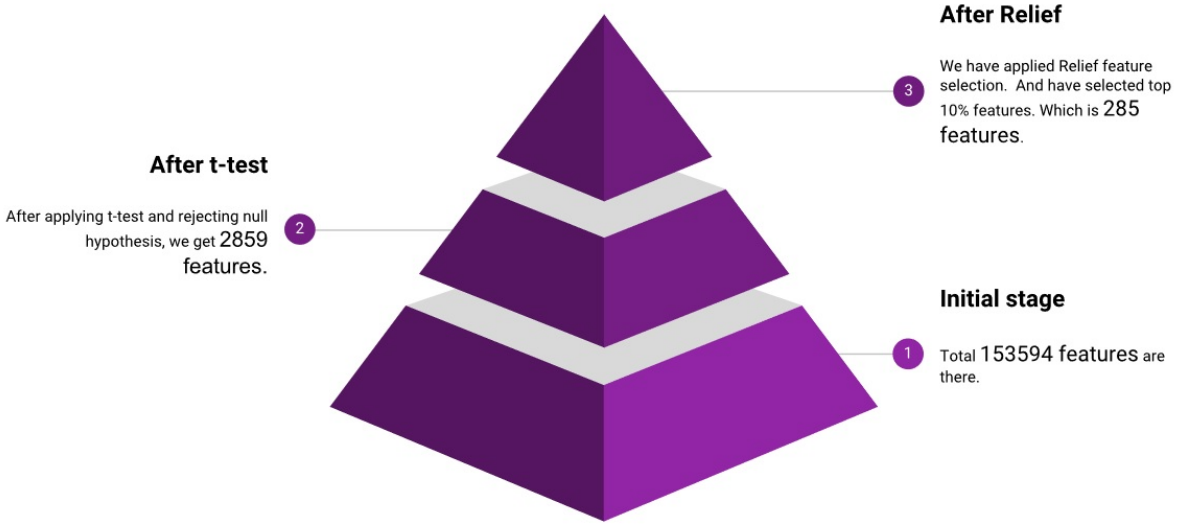

FIGURE 4.2. Number of features after each feature selection

20% i.e. 570 features, top 30% i.e. 855 features, top 40% i.e. 1140 features.

LooCV- The full form is Leave one out Cross Validation. That means when training and testing the model, we keep 1 subject out for testing

purpose and take rest for training the model. When the model is trained, we test out new model with the testing subject to check whether our model is working good or not. This process is done for each subject making test subject and rest train subject. After that the mean is taken to calculate the final accuracy.

60/40- In this method, the dataset is divided into two parts randomly. 40% of the data goes to one part and rest 60% of the data goes to another part. Those 60% data are used for training the model and 40% data are used to evaluate the model.

80/20- In this method, the dataset is divided into two parts randomly. 80% of data goes to one part and rest to another. Those 80% data are used to train the model and 20% data are used for testing purpose.

4.2 Results of ELM

This is done taking top 10% features. We have randomly taken hidden neuron numbers in between 100 and 3000 with sigmoid activation function. We have done this 250 times. The graph shows the number of hidden neurons in the X-axis and accuracy on the Y-axis. The graph looks like fig 4.3. Now we have increased the number of hidden neurons. Now we are taking random hidden neuron numbers from 100 to 5000. The graph looks like fig 4.4. Now we have increased the number of hidden neurons. Now we are taking random hidden neuron numbers from 100 to 7000. The graph looks like fig 4.5

So, we can see that as we increase the number of hidden neurons, the accuracy increases. But after a certain number, it stabilizes. So, after around 4500 hidden neurons we are able to get 98.16% accuracy. This experiment is done in LooCV method.
We have done the experiment in different ways too. We have divided the dataset into 60-40. That means we have taken 60% of the total subjects

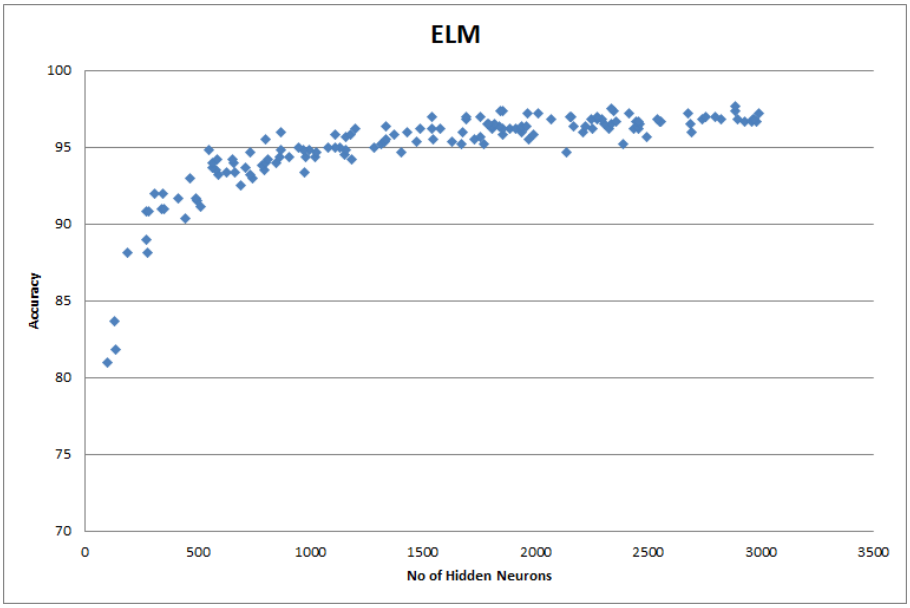

FIGURE 4.3. Graph of accuracy and number of hidden neurons for ELM

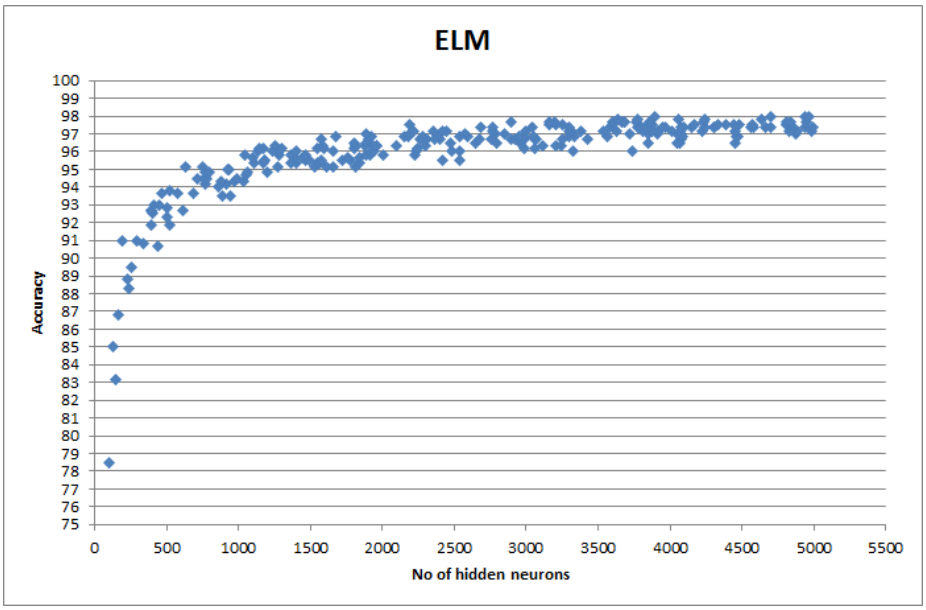

FIGURE 4.4. Graph of accuracy and number of hidden neurons for ELM

randomly and marked them as a training set. The other 40% subjects are marked as the testing set. We have trained our model using that training

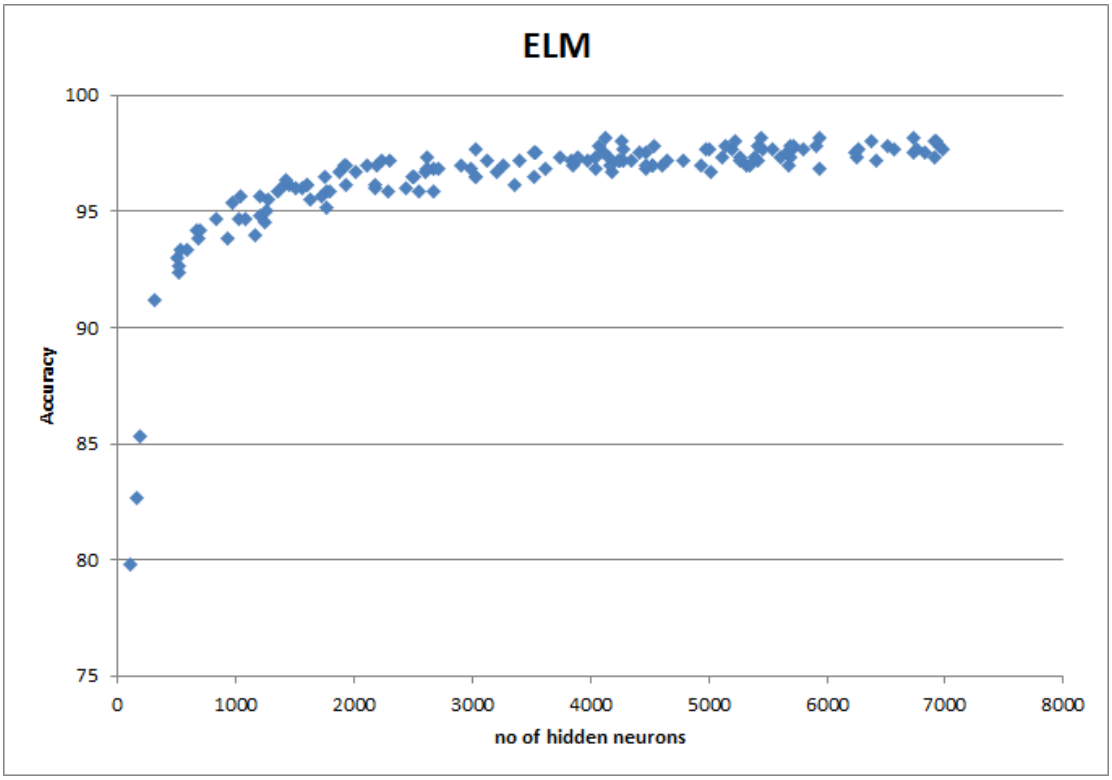

FIGURE 4.5. Graph of accuracy and number of hidden neurons for ELM

set and then have tested on the testing set to calculate the accuracy. This whole process was repeated 20 times and at last, we have taken the mean of all those accuracies to get the final accuracy. We are able to get 84.9% accuracy in this method using 4122 hidden neurons and sigmoid activation function. The confusion matrix looks like fig 4.6. We have done the same in 80-20 division also. Again after 20 times repeating the process and taking the mean we have got accuracy of 85.25%. The confusion matrix looks like fig 4.7.

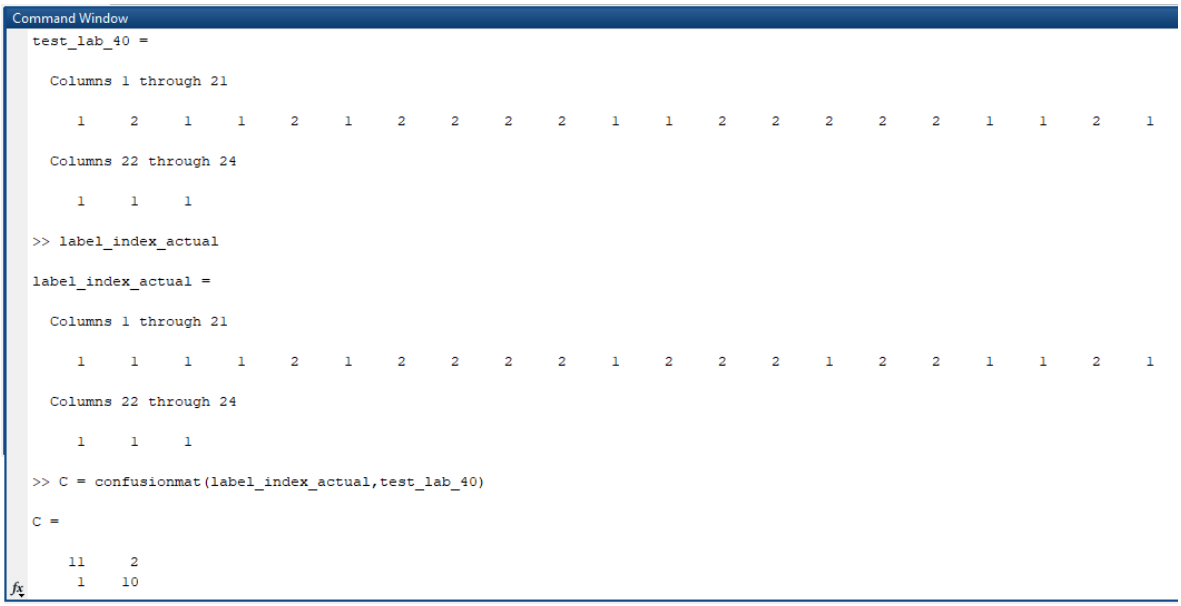

FIGURE 4.6. Confusion Matrix

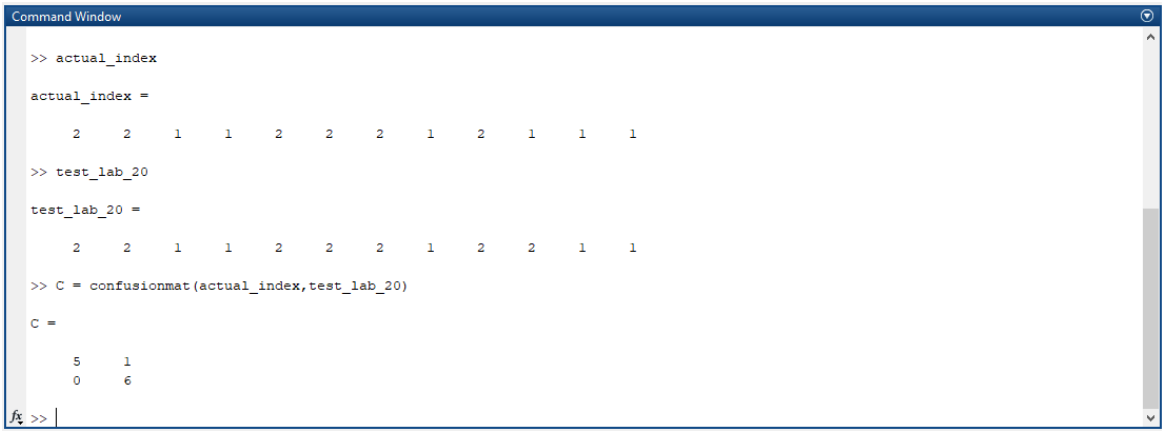

FIGURE 4.7. Confusion Matrix

4.3 Results of Kernel ELM

We have applied Karnel ELM with 500 hidden neurons and with 'RBF kernel', for which the accuracy we got is 80%. Also with same attributes and 'linear kernel' gives accuracy of 71.67%. This process is done in LooCV method i.e. every time we take 59 subjects as training and 1 subject as

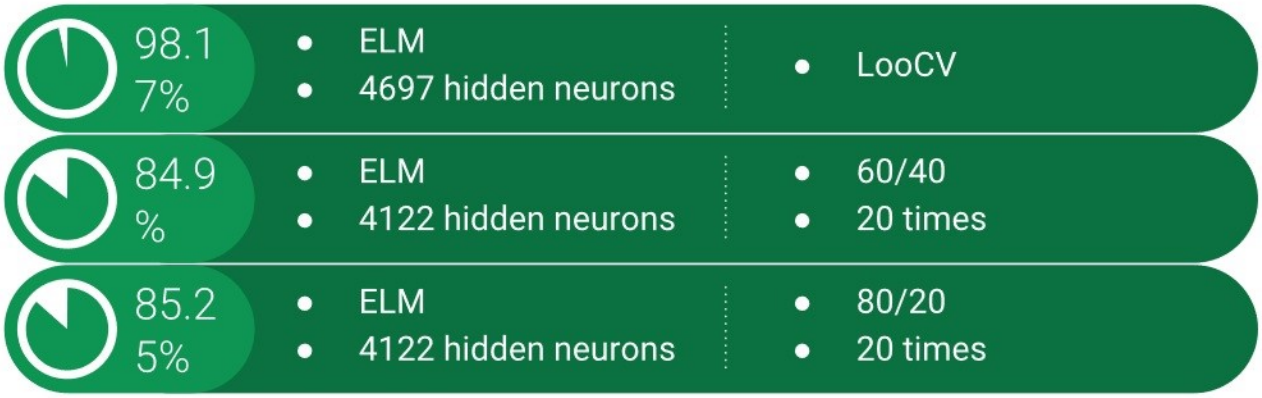

FIGURE 4.8. Accuracy table of ELM

	60/40	**80/20**
Linear Kernel	81.25%	79.17%
RBF Kernel	50.25%	49.58%

Table 4.1: Results of Kernel ELM

testing. This process is done 60 times and the mean is taken for final accuracy. The accuracy stays same with increase of hidden neurons.
If we apply the same Kernel ELM with 500 hidden neurons and with 'lin kernel' in 60-40 method i.e. 60% of data is used for training and rest 40% of data is used for testing, then we get accuracy of 81.25%. This is done 20 times and at the end the mean of this 20 runs is taken which is 81.25%. The same process with 'RBF kernel' gives 50.25%.
If we apply the same Kernel ELM with 500 hidden neurons and with 'lin kernel' in 80-20 method i.e. 80% of data is used for training and rest 20% of data is used for testing, then we get accuracy of 79.17%. This is done 20 times and at the end the mean of this 20 runs is taken which is 79.17%. The same process with 'RBF kernel' gives 49.58%.

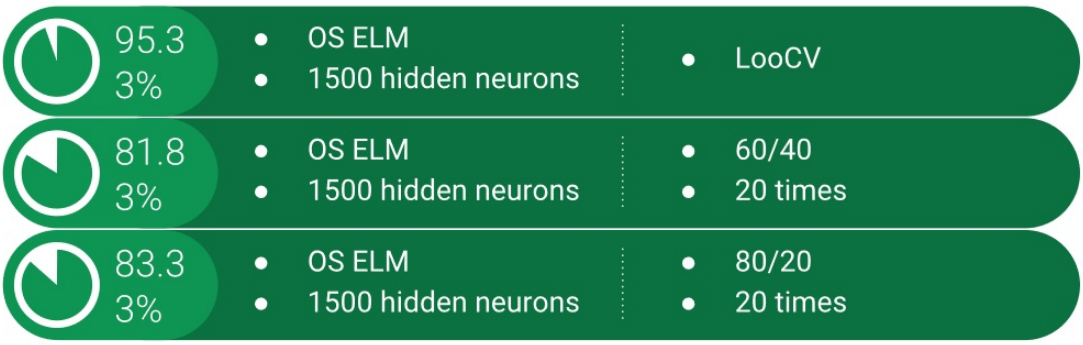

FIGURE 4.9. Accuracy table of OSELM

4.4 Results of OS ELM

In online sequential ELM, we have taken 1500 hidden neurons. In LooCV we have got 95.33% accuracy. The same process has been done in 60/40 method. After doing 60/40 20 times and taking the mean we get the accuracy of 81.33%. Similarly, 80/20 after running 20 times and taking mean the accuracy comes out to be 83.33%.

4.5 Comparative Results

Here is the comparative graph between 5 classifiers(Figure 4.10). Those classifiers are KNN(k nearest neighbor), SVM(support vector machine), ELM(extreme learning machine), OS-ELM(online sequential extreme learning machine), Kernel ELM. As we can see the ELM classifier gives the best result, followed by OS ELM, SVM, Kernel ELM, KNN.

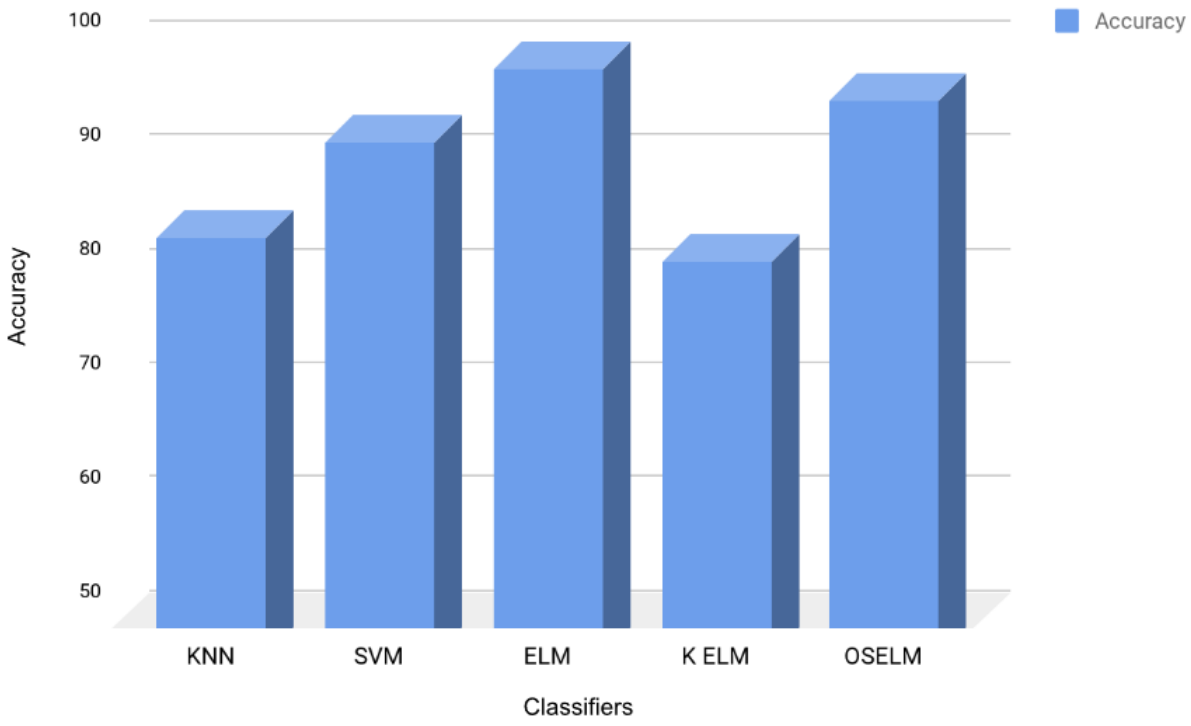

FIGURE 4.10. Comparison between different classifiers

4.6 Brain affected regions

We have selected features using different feature selection algorithms. Now when we track back to those regions, we can see mainly affected brain regions of brain for schizophrenia disease affected persons.

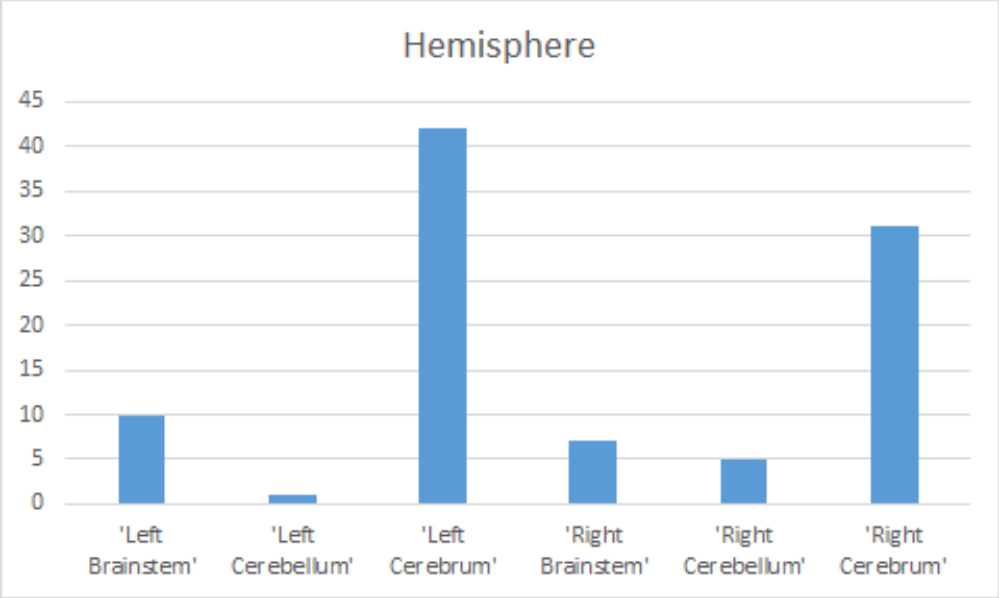

FIGURE 4.11. Percentage wise distribution of affected voxels covering hemisphere regions

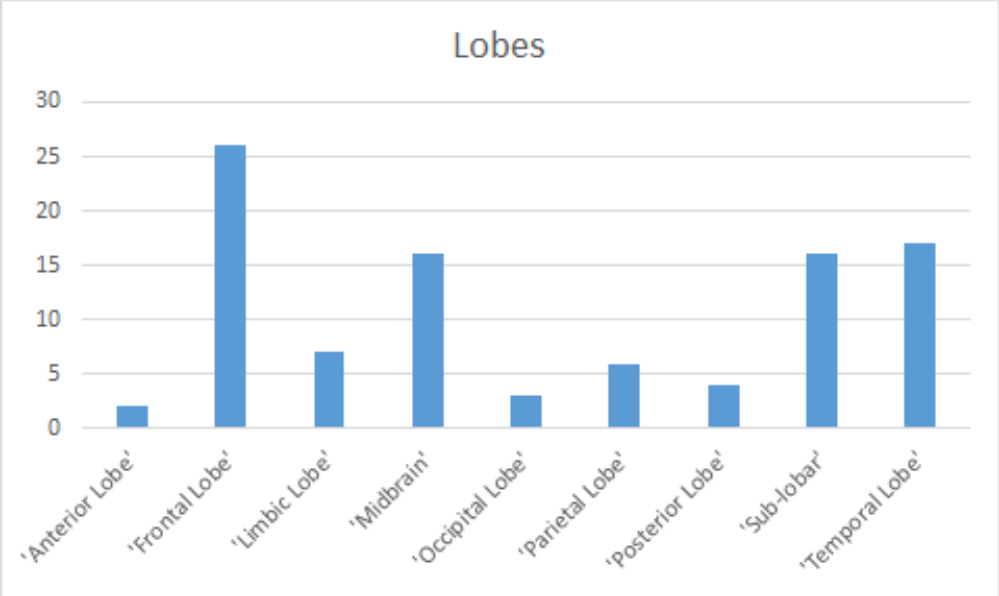

FIGURE 4.12. Percentage wise distribution of affected voxels covering the lobes

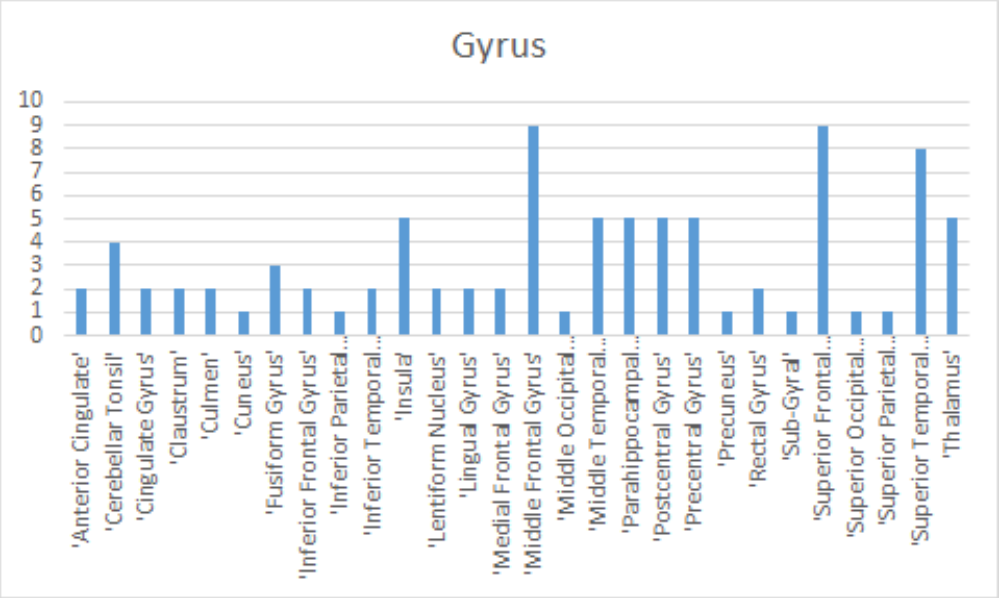

FIGURE 4.13. Percentage wise distribution of affected voxels covering gyral regions

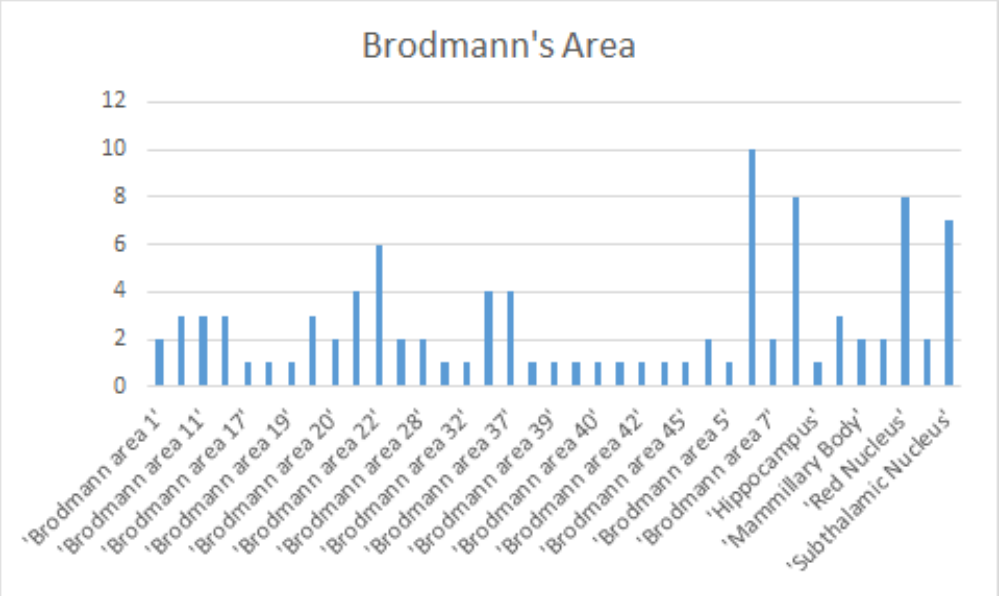

FIGURE 4.14. Percentage wise distribution of affected voxels covering Brodmann's areas

BIBLIOGRAPHY

[1] M. ALEXANDRESAVIO, *Local activity features for computer aided diagnosis of schizophrenia on resting-state fmri*, Neurocomputing.

[2] C. S. C. D. M. B. V. A. S. J. D. C. AVRAM J. HOLMESA, ANGUS MACDONALD, *Prefrontal functioning during context processing in schizophrenia and major depression: An event-related fmri study*, elsevier, 76 (2005), p. 199– 206.

[3] L. L. R. V. A. S. CAMERON S. CARTER, ANGUS W. MACDONALD, *Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: An event-related fmri study*, Am J Psychiatry, 158 (2001), pp. 1423–1428.

[4] M. G. DARYA CHYZHYKA, ALEXANDRE SAVIO, *Computer aided diagnosis of schizophrenia on resting state fmri data by ensembles of elm*, Elsevier Ltd., 68 (2015), pp. 23–33.

[5] A. F. M. G. S. A. M. GRAZIELLA ORRÙA, WILLIAM PETTERSSONYEOA, *Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review*, Neuroscience and Biobehavioral Reviews, 36 (2012), p. 1140–1152.

[6] Y. L. D. H. HUI SHEN, LUBIN WANG, *Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fmri*, NeuroImage, 49 (2010), p. 3110–3121.

[7] L. A. R. KENJI KIRA, *A practical approach to feature selection*, (1992).

[8] B. B. J. G. R. C. P. O. R. P. B. B. D. K. D. T. R. I. G. S. E. T. C. J. J. M.M. MACHULDA, H.A. WARD, *Comparison of memory fmri response among normal, mci, and alzheimer's patients*, NEUROLOGY, 61 (2003), p. 500–506.

[9] V. A. M. N. C. A. J. L. K. A. K. G. D. P. V. D. C. OGUZ DEMIRCI, VINCENT P. CLARK, *A review of challenges in the use of fmri for disease classification / characterization and a projection pursuit application from a multi-site fmri schizophrenia study*, Springer Science + Business Media, 2 (2008), pp. 207–226.

[10] V. D. C. OGUZ DEMIRCI, VINCENT P. CLARK, *A projection pursuit algorithm to classify individuals using fmri data: Application to schizophrenia*, NeuroImage, 39 (2008), p. 1774–1782.

[11] G. D. P. VINCE D. CALHOUN, PAUL K. MACIEJEWSKI AND K. A. KIEHL, *Temporal lobe and default hemodynamic brain modes discriminate between schizophrenia and bipolar disorder*, Human Brain Mapping, 29 (2008), p. 1265–1275.

[12] ——, *Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder*, Human Brain Mapping, 29 (2008), p. 1265–1275.