Chapter 4

SENTENCE POLARITY DETECTION

Reading the state of the brain was always a fascinating task for humanity. The advent of new non-invasive techniques like fMRI (Functional Magnetic Resonance Imaging), PET (Positron Emission Tomography) scans have made it possible to record the state of the brain at the same time; healthy subjects are engaged in certain specified activities. fMRI study of syntax in the brain and especially sentence polarity detection in the brain has been evolved as a charming area. In this Chapter, we have analysed the fMRI data corresponding to the brain's affirmative and negative sentence processing. Using a greedy stepwise correlation-based feature selection technique and random forest classification technique, our model can classify the cognitive state in sentence polarity detection task with, on average, 95.41% accuracy. We have also analysed the category-specific selected feature voxel set to determine the sentence polarity in the brain. Our result shows that CALC, RDLPFC, LDLPFC are highly contributing areas in feature selection. In contrast, RPPREC, RSGA, and RFEF add little to polarity detection.

4.1 INTRODUCTION

Making computational models which can recognize the cognitive state [230] using brain data was always an exciting assignment for humanity. The discovery of new non-invasive techniques like fMRI, PET scan has made it possible to record the state of the brain while healthy subjects are involved in doing specific specified actions. fMRI (Functional magnetic resonance imaging) is a technique that captures and represents a three-dimensional image of brain activity during the study. The fMRI scanner displays the fMRI signals values in all the points in three-dimensional grids called voxel in analogues to the pixel in two-dimensional images [231]. fMRI data has been utilised to find out the activation pattern and the localization of the activated voxels in the brain while the subject is engaged in some specified tasks. Many of the research work published in the literature focused on finding specific brain regions that are activated during the performance of particular activity like viewing some image, processing word, sentence, etc. Syntactic comprehension is the central aspect of human language, and syntax is processed in the brain differently than other features of language –morphology, semantics, etc. [232]. The study and the representation of the syntax (sentence structure) in the brain based on neuroimaging data are enriched in new findings. The study of the syntactic brain grounded on fMRI data can be viewed in five types of sentence arrangements in literature. A) Syntactic violations B) Complex vs. simple sentences C) Sentence vs. word list D) sentences containing pseudo words E) sentences having separate agreements and types. In case of syntactic violation, in which the sentences are syntactically correct but semantically incorrect or having some grammatical error or spelling error caused increased activation in areas that are involved in syntactic processing due to disruption in agreement checking and structure building which results in more attentive behaviour [233]. More superior frontal activity (BA 6, 8) is found for syntactic error compared to any other errors. In grammatical vs. ungrammatical sentences, both types of sentences were presented to the subject, and corresponding fMRI activation was recorded. Syntactically correct and incorrect sentences have different activation patterns and areas of activation in the brain [116]. For complex and simple sentences, the complex sentence involves more syntactic operations, i.e., more reconstruction and canonical word ordering hence activates more area than simple sentences [234]. The enhanced activation in complex sentences lies in Broca's area (BA 44, 45) extended to BA 47, 6, and 9 [235]. Some additional activation is found in left bilateral superior and middle temporal gyri (BA21, 22), left angular supramarginal gyri (BA 39, 40) and cingulate gyrus (BA23,24,31,32). In the case of sentence vs. word list where a comparison is made between sentence, i.e., syntactic structure and a list of unrelated words, an increased activation is observed for sentences in the anterior part of the temporal lobe, especially temporal pole (BA 38). In addition to this, superior and middle temporal gyri (BA 22 and 21) are also found activated [236]. The comparison of the simple sentence with the pseudoword containing sentences, i.e., sentences that are syntactically correct but consisting of meaningless words (pseudo-words), yielded activation in the posterior superior temporal sulcus (BA 22,41/42). Additionally, anterior superior temporal sulcus (BA 38,22) was also found activated. However, syntactically and semantically correct sentences are studied by altering its structure and agreements. Lots of research has been done to date, where different forms of sentences are reviewed. In [237], authors have demonstrated the mechanism of case processing in our brain, and the result found shows that the processing of the genitive case activates the left inferior frontal gyrus and posterior part of the middle temporal gyrus more than the nominative case and accusative case for English native speakers. The negative sentences have more syntactic structure than affirmative sentences since they contain other syntactic entities for negation [238]. Different negative sentences have diverse representations in the brain, depending on whether these include bipolar predicate or contradictory predicate [239]. Processing of negative sentences is assumed to be a two-step process in which, at first, the affirmative form is processed. Then the negative version of it is represented in the brain, whereas affirmative sentences are directly processed in the brain [240]. The additional syntactic transformation reflects greater cortical activation in the case of negative sentences. Higher activation in the left posterior temporal gyrus and the bilateral parietal brain is observed in negative sentences compared to its affirmative counterpart [241]. In the Japanese-English sequential bilingual paradigm study with target-probe matching task, a significant activation in the left temporal and left precentral gyrus was observed for negative sentences in English, which is the second language for the participants [242]. The study of action-related and abstract sentence polarity sentences has revealed a dynamic mental simulation model of negative sentence processing. Increased left hemisphere perisylvian and parietal activation was observed for abstract negative sentences, but a partial deactivation was found in action-related negative sentences in the left pallidum [243]. In [244] for the Danish language, increased activation in the left premotor cortex (BA6) for negative sentences was found. This study reveals that there are three critical cortical centres for the processing of negative sentences- left premotor cortex (BA6) for sentence structure, bilateral inferior parietal (SMG, BA 40) for semantic processing, and left inferior frontal gyrus dedicated for computation of syntactic complexity. The authors examined the neural correlates of the differences between negative and affirmative sentences using functional magnetic resonance imaging (fMRI), and results found showed enlarged activation in left premotor cortex from negation companionable with rule-governed memory processing and increased activation in the right supramarginal gyrus from affirmation, compatible with semantic processing. In [245] in fMRI study of sentences having single and double negation for the German language, it is found that left pars triangularis (BA 45) left pars opercularis (BA44) left supplementary motor area (SMA(BA6)) and left superior temporal gyrus (STG(BA42)) are functionally associated in the processing of main clause negation. IFG (inferior frontal gyrus) plays the most crucial role in the coordination of the cortical areas responsible for language processing and logical reasoning in the interpretation of negative sentences. In [246] authors have examined fMRI data for activation pattern of the brain in the processing of negative/affirmative sentences of Hindi and found that common cortical region involved in processing includes-bilateral inferior frontal gyrus (IFG), left parietal cortex (BA40), left fusiform (BA37), bilateral supplementary motor area (SMA (BA6)), bilateral temporal gyrus (BA21) and bilateral occipital area (BA17/18). In addition to this, the study of fMRI data shows that the anterior temporal pole is dedicated to the processing of negative sentences. In this chapter, we have analysed the star plus data-set [247] available online to classify the neural activity of the brain corresponding to the polarity of sentences. The star-plus data set is used in different types of analysis, like the cognitive state of the brain [248], MVPA pattern analysis of brain signals [249], deep knowledge representations [250], and many more. We have extracted the fMRI data dedicated to sentence processing from it, and using a greedy stepwise feature selection approach contributing feature value data is selected. From this, applying a random forest classifier to classify the brain state, our model can recognize with 95% accuracy when a person is reading a negative sentence or an affirmative sentence. This proposed method aims to recognize the state of the brain, whether it is processing a negative sentence or affirmative sentence by analysing the fMRI signal of the brain. The result analysis shows that the model can recognize the brain state with more than 95% accuracy.

4.2 METHODS AND MODELS

4.2.1 DATA-SET DESCRIPTION

The data set used in this Chapter is known as the Star Plus data set, which was collected by Marcel Just and his colleagues at Carnegie Mellon University. The fMRI signals were collected over a grid of 64*64*8 voxels throughout the experiment. The data was collected from six healthy subjects. Each subject was performing a sequence of activities during the entire session of fMRI recording. Activities include showing a picture and descriptions of it in sequence and then pressing a button of yes or no depending upon whether the sentence correctly describes the picture or not. This event-related experiment consists of trials. One trial of experiments lasts for 27 seconds. At every 500ms interval, fMRI data was recorded for the whole-brain; hence a total of 54 images of the brain were recorded in a specific trial of the experiment. At the start of the experiment, a picture or sentence was presented on the screen for 4 sec, and then after another 4sec, a blank screen was shown. After that, a picture or sentence was shown again for 4 sec, in which a button was to be pressed for yes or no whether the sentence correctly describes the picture or not. If the first stimulus is shown as a picture, then the second item displayed was a sentence and vice versa. The picture first and sentence after was termed as PS data set, and in the case where the sentence was the first stimulus, the data was termed as SP data set. In half of the trial, the picture was presented first, and the remaining half trial

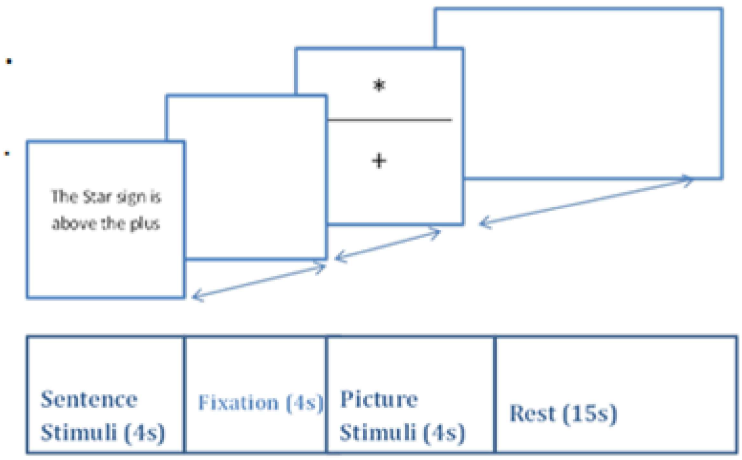

FIGURE 4.1: Trial Representation

consists of presenting sentences as the first stimulus. The picture presented in this experiment were geometrical arrangements of \$, *, and +. Whereas the sentences presented were a description of a picture and half of the sentences were affirmative sentences and the remaining half were negative sentences. The sentences were just a description of the picture like ”it is true that the dollar is above the star” or ”it is not true that the star is above the plus.” A total of 40 trials were presented to each subject. Each trial consists of around 5000 voxels for one image, so a total of 270000 voxels for one trial of the experiment. The whole brain is divided into 25 ROIs (region of interests) into which different voxel activity patterns reside.

4.2.2 MODEL

The stepwise realisation of the proposed approach can be viewed in the following flow diagram in fig 4.2

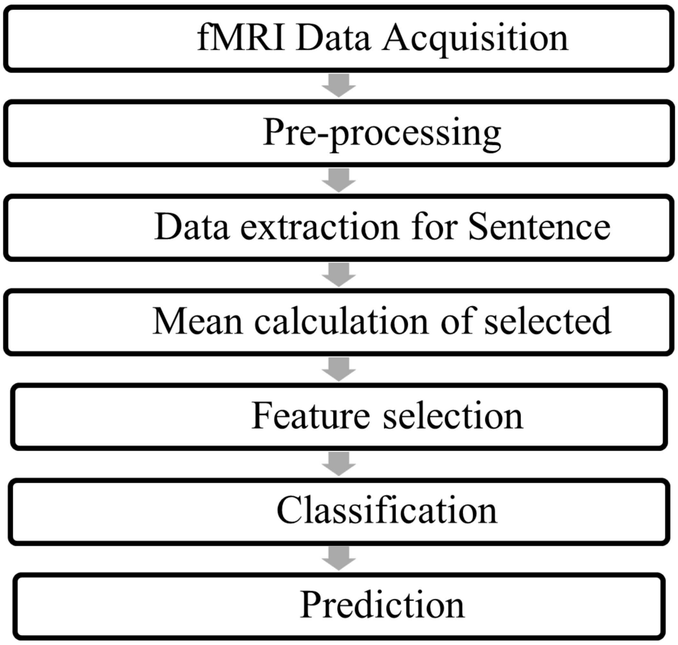

FIGURE 4.2: Proposed Model

4.2.2.1 fMRI DATA ACQUISITION

T2 weighted -fMRI data were collected from 6 healthy people at the interval of every 500ms. In this experiment, a 3.0 T GE Signa scanner was used, having TE=18ms, TR=500ms, and a flip angle of 500. The resulting images captured approximately

5000 voxels per subject in 8 oblique axial slices in two separate non-contiguous four-slice volumes [251]. An in-plane resolution of 3.125 mm and a slice thickness of 3.2 mm was used, resulting in approximately 5000 voxels per participant.

4.2.2.2 PRE-PROCESSING

fMRI images were pre-processed using the FIASCO program [252] to remove artefacts, i.e., head motion and signal drift, and correcting slice timing. Talairach coordinate was used for the normalisation of data. Twenty-five anatomical regions of interest (ROIs). The regions which were selected after pre-processing which includes calcarine sulcus (CALC), left intraparietal sulcus (LIPS), left dorsolateral prefrontal cortex (LDLPFC)left inferior frontal gyrus (LIFG) left opercularis (LOPER), left and right inferior temporal lobule (LIT, RIT), left and right posterior precentral sulcus (LPPREC, RPPREC), left frontal eye fields (LFEF), left inferior parietal lobule (LIPL), left temporal (LT) lobe, left and right triangularis (LTRIA, RTRIA), right temporal lobe(RT), left and right supramarginal gyrus (LSGA, RSGA), left superior parietal lobule(LSPL), right superior parietal lobule (RSPL)),) right inferior parietal lobule(RIPL), right frontal eye fields (RFEF), right opercularis (ROPER), right dorsolateral prefrontal cortex(RDLPFC), right intraparietal sulcus (RIPS), and supplementary motor areas (SMA).

4.2.2.3 DATA EXTRACTION FOR SENTENCE PROCESSING

The data for sentence processing is extracted from the Star Plus data set. The Star Plus data set is composed of two types of data. In one type sentence was shown as the first stimuli and 8 sec later a picture was shown named as SP dataset. In another type picture the first and later sentence is shown. The fMRI signals corresponding to the sentence are extracted from the data set to analyse the stimuli for sentence processing in the brain. The Star-Plus data set is assumed to be composed of two data sets- the PS data set and the SP data set. In the PS data set where the picture was first stimuli, the onset of the sentence appears at t=8 sec on the screen and lasts for 4 seconds. In the SP dataset sentence was the first stimulus, and it appears on the screen at t=0 and lasts for 4 seconds after that 4 sec of rest period was there. Hence for SP data set from t=0 to t=8, a total of 16 brain images were extracted. Similarly, in the case of the PS data set, images were extracted from t=8 to t=16, again a total of 16 images. Although stimulus for the sentence was presented only for 4 sec, we have taken an 8-sec interval of fMRI data to train our classifier because fMRI BOLD signals often extend for 9-12 sec beyond the neural activity of interest. These images are further used by feature selection entities to find prominent feature stimuli sets.

4.2.2.4 MEAN CALCULATION

The extracted data for classifier training takes the form-

$$fMRI - sequence(t, t+8) - -- > Affirmative/Negative\ sentence \qquad (4.1)$$

The data set so extracted is still an enormous size to analyse. The resulting matrix for each subject obtained to be 16*5000 for one trial of SP or PS data. We calculated the mean of all the 16 images for all the feature vector points. This reduces the size of data in the processing one specific sentence to the matrix to 1*5000 from 16*5000 in one trial. An array of 40*5000 is thus obtained for a single subject in the processing of all the 40 trials for the processing of all the negative and affirmative sentences. The mean is calculated over entire column entries by the following formula-

$$\forall J\ J\epsilon Column, \sum_{i=1}^{n} x_{ij}/n\ Where\ x_{ij}\ is\ value\ at\ ith\ row\ and\ jth\ column \qquad (4.2)$$

4.2.2.5 FEATURE SELECTIONS

• INFO GAIN-After the mean calculation, the prominent feature set is obtained using info-gain feature selection [253],[254]. It measures the contribution of each feature in decreasing the overall entropy. The Entropy, H(X), is defined as follows:

$$H\left(X\right)=-\sum\left(P_{i}*log_{2}(P_{i})\right) \tag{4.3}$$

Where Pi-> is the probability of class i in the dataset, a good attribute is an attribute that contains the most information, i.e., reduces most of the entropy. The contributing feature set is fed to the classifier to classify the brain state using the k-NN classifier [255].

• ENTROPY BASED FEATURE SELECTION- For classification, the feature which is highly correlated with the class and having minimal correlation with the other feature is selected considering correlation as goodness measure in feature selection [256]. This feature selection technique works in two steps- first the features are selected relevant to the class and second, iteratively eliminate other features based on the selected feature. Feature which is relevant to the class concept and not redundant to other relevant feature assumed to be good. Correlation measure based on information theoretical concept of entropy for variable x is given by

$$\mathbf{H}(\mathbf{X})=-\sum_{i=1}^{m}P(x_{i})log_{2}(\mathbf{P}(x_{i})) \tag{4.4}$$

The entropy of after observing the variable is represented by

$$H(X/Y)=-\sum_{j=1}^{N}P(y_{j})\sum_{i=1}^{M}P(x_{i}/y_{j})log_{2}(P(x_{i}/y_{j})) \tag{4.5}$$

The range of j varies from 1 to N, where N is number of voxels(n). Where P(x_i) *are the prior probability for all values of* x *and* $P(x_i/y_i)$ *is the posterior probabilities of* X *given the values of* Y

The information gain is the decrease in entropy of X provided by Y and is given by amount by

$$IG(X/Y) = H(X) - H(X/Y) \tag{4.6}$$

Correlation being biased in favor of features having more value symmetry becomes a measure of the correlation between features. Symmetrical uncertainty is represented as

$$SU(X,Y) = 2\left[\frac{IG(X/Y)}{H(X)+H(Y)}\right] \tag{4.7}$$

Taking symmetrical uncertainty as a goodness measure, relevant features are selected. The range of SU lies between 0 and 1. After finding the values of SU for all features, we have chosen the top dominating voxels based on the threshold value of SU. For data set S containing N features and C classes, symmetrical uncertainty is denoted as SUic for correlation between Fi and class C. then

$$\forall\ F_i \in\ S`, 1 \leq\ i \leq\ N, SU_i \geq \delta, (\delta \rightarrow\ threshold) \tag{4.8}$$

The procedure recursively eliminates the irrelevant feature from the fMRI data and selects only correlated ones.

- CFS Feature Selection-Selecting the feature set comprises three methods, namely attribute evaluator, search method, and attribute selection mode. The goal of the feature selection method is to identify the feature subset that is minimal in size and collectively optimally predictive by removing irrelevant as well as redundant features. An attribute evaluator is a correlation-based feature selector that selects the subset of highly correlated features but has low intra-correlation. CFS as a filter algorithm makes a ranking of feature subset based on the correlation heuristic evaluation function. Irrelevant features are ignored as they are less correlated with the class, while the redundant feature is screened out being positively correlated with one or more features. CFCs feature subset evaluation function [257] is defined as

$$M_s = \frac{kr_{cf}}{\sqrt{k + k\left(k-1\right)r_{ff}}} \tag{4.9}$$

$\mathrm{M}_s = Heuristic\ merit\ of\ a\ feature\ subset\ S\ containing\ k\ features$

$\mathrm{r}_{cf} = mean\ feature\ class\ correlation\ and$

$\mathrm{r}_{ff} = Average\ feature - feature\ inter - correlation$

The numerator of the equation denotes how predictive of the class a set of features are, whereas the denominator indicates the measure of redundancy among the features. The search method is a greedy stepwise forward-backward algorithm [258] that performs a forward or backward search through the space of the attribute subset. The forward iteration, the feature, which provides the largest increase in terms of predictive performance, is selected. The forward phase ends when no feature further improves performance or a maximum number of selected features has been reached. In the backward phase, the feature which does not reduce the performance is removed, and the backward phase stops when no feature is being removed without reducing the performance.

Attributes are selected over the entire data set to make selection justifiable. After the attribute selection phase is over-classification is done. The overall working of CFS feature selection can be understood by the following figure.

Data is divided into two parts: training data and testing data. Training data is fed to the filter-based correlation module where, using the heuristic equation, non-correlated and redundant features are removed. The dimensionality reduction module reduces the dimension of data. Here forward, data is fed to the machine learning module for classification.

Algorithm 4 Forward–Backward Selection

function FBSELECTION(I, TD, L)
    Input: InputData I, TargetData TD, Significance Level L
    Output: Selected Features **SF**
    SF $\leftarrow \phi$
    // Empty. Selected features are initially
    // in forwarding phase. Iteration is done until //
    // no more features can be selected //
    while SF*changes* **do**
        // conditional on SF with minimum p-value, V^* Identified //
        $V* \leftarrow argmin\ PVALUE(TD, V|SF)$
        $V \epsilon V_D \setminus S$
        // if conditionally dependent with TD given SF, Select V^*
        if $PVALUE(TD, V^*|SF)L$ **then**
            $SF \leftarrow$ SF $\cup V^*$
        end if
    end while
    //In backward phase, until no more features can be removed from S,
    // Iteration is performed
    while SF*changes* **do**
        //Identify V^* with maximum p-value conditional on SV^*
        $V^* \leftarrow argmax\ PVALUE(TD, V|\mathbf{SF} \setminus V)$
        $V \epsilon S$
        //Remove V^* if conditionally independent with TD given $\mathbf{SF} \setminus V^*$
        if $PVALUE(TD, V^*|\mathbf{SF} \setminus V^*) > L$ **then**
            SF$\leftarrow$ **SF** $\setminus V^*$
        end if
    end while
    return SF
end function

4.2.2.6 CLASSIFICATION

k-NN

k-NN is a non-parametric learning, lazy algorithm that uses feature similarity to predict the value for new data points [259]. The unique data point is assigned a value based on its similarity with the data points in training data. The steps of the k-NN algorithm are as follows-

1. Load the training as well as test data
2. Choose the value of K- the number of neighbours and initialise it.

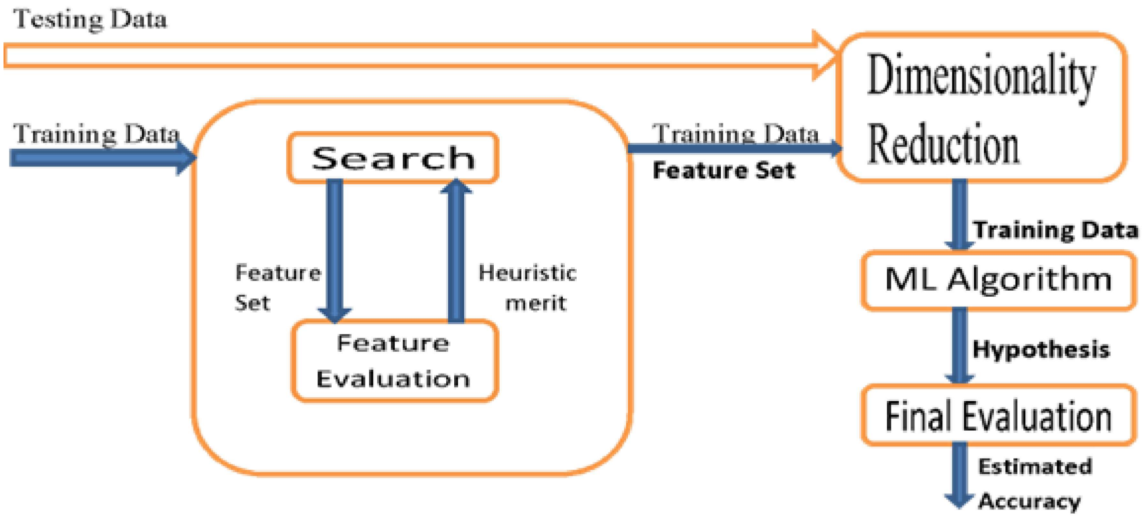

FIGURE 4.3: CFS Feature Selection

3. For each data point in test data

- a. Calculate the Euclidean distance between the test data and each row of training data.
- b. Add the index and the distance of the test data to an ordered collection

4. Sort the ordered collection of indices and distances in ascending order.
5. Pick up the top K rows from the sorted collection
6. Get the labels of the selected K entries
7. Return the mode of the K labels

MULTILAYER PERCEPTRON[1]

A multilayer perceptron (MLP) [260], is a class of feed forward artificial neural network. MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that

[1]http://deeplearning.net/tutorial/mlp.html

uses a nonlinear activation function. MLP utilises a supervised learning technique called back propagation [2] for training.

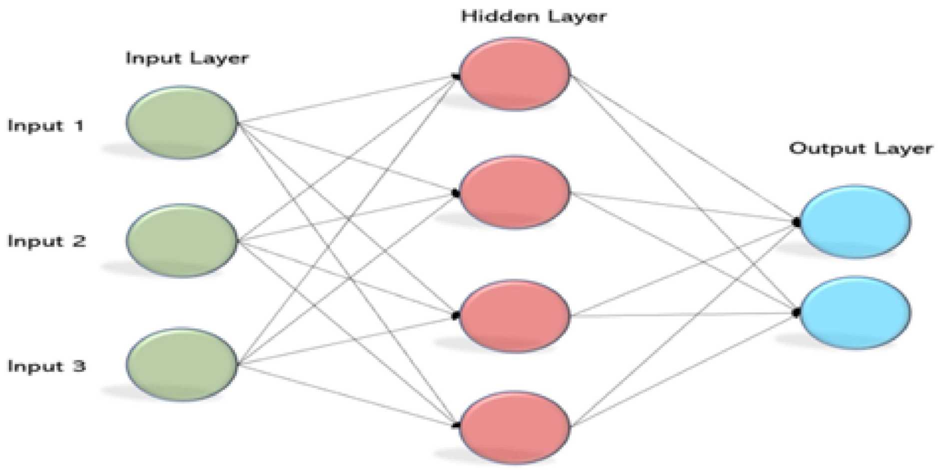

FIGURE 4.4: Multilayer Perceptron

Single-hidden-layer MLP function can be seen as f:$R^D \rightarrow R^L$,

where the size of input vector xis D,

and the size of the output vector f(x) is L , The matrix notation is given as

f(x)=G($b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x))$)

with $b^{(1)}, b^{(2)} - - > biasvectors;$

$W^{(1)}, W^{(2)} - - > weightmatrices, and,$

G and s are activation functions. The hidden layer is

h(x)=$\varphi(x) = s(b^{(1)} + W^{(1)}x)$

Where,$W^{(1)} \in R^{D*D_h}$

is the weight matrix connecting the input vector to the hidden layer. Each column $W_i^{(1)}$

[2] https://becominghuman.ai/multi-layer-perceptron-mlp-models-on-real-world-banking-data-f6dd3d7e998f

represents the weights from the input units to the ith hidden unit

The choice for s is given by tanh(a)=(e$^{a}-e^{-a})/(e^{a}+e^{-a})$

The output vector is then obtained as: O(x)=G(b$^{(2)}+W^{(2)}h(x))$

RANDOM FOREST

We have implemented a random forest classification technique for the analysis and classification of the dataset [138]. Random Forest Classifier is an ensemble algorithm. More than one algorithm of the same or different kinds is combined at the same time in the case of ensemble algorithms in classification tasks. A set of the decision tree is created by the Random Forest classifier from a randomly selected subset of the training set. The votes from different decision trees are aggregated to decide the final class of the test object. The random forest algorithm steps

1. Random samples are selected from the dataset.
2. A decision tree is constructed for every sample, and prediction results from every decision tree is obtained.
3. For every predicted result, voting is done
4. At last, select the most voted prediction result as the final prediction result.

The problem of overfitting is solved by random forest as it combines or averages the results of different decision trees. The result obtained for the large dataset is more accurate than a single decision tree. Random forests are very flexible and possess very high accuracy. Scaling of data is not required in the Random Forest algorithm. It maintains good accuracy even after providing data without scaling. A Random Forest algorithm maintains good accuracy; even a large proportion of the data is missing.

4.3 RESULT ANALYSIS AND DISCUSSION

The data was collected from six persons, and each person has its own identifier id or subject id. For all six persons, their identifiers are the following: Subject 04799, Subject 04820, Subject 04847, Subject 05675, Subject 05680, and Subject 05710. The data set comprises fMRI signals distributed over 24 different anatomical regions of the brain. The analysis of the data set reveals that for the same task, the numbers of voxels involved differ not only anatomical area-wise but subject-wise also. The region-wise distribution of voxel activities is presented in table 4.1 which shows the distribution of signal data into different regions of the brain.

The data corresponding to the sentence processing is extracted according to the eq.1 from the star plus dataset. The mean of the extracted data is calculated by eq.2.and this data is further analysed to study the state of the brain corresponding to the sentence polarity task. Employing stepwise greedy correlation-based feature selection method on data, we obtained a highly contributing feature set removing the redundant features. Fig.4.5 shows the statistical distribution of selected features values for subject 3, and from this representation, it is clear that selected feature vectors are distributed in nature and not skewed. The distribution is not right-skewed or left-skewed, and most of the selected feature vectors are not redundant and almost different from each other. These distributed, non-skewed, and less redundant selected features justify that our feature selection technique is correct and select prominent features only. The blue colour denotes the affirmative data, whereas red meant negative data values. We have plotted the distribution of feature values on the x-axis, which is from min to max values. The y-axis represents the number of samples. Corner value is simply the feature number, and values on the x-axis represent max to min. Y-axis represents the number of samples.

The region-wise frequency of selected feature voxel does not provide any specific conclusion regarding which prominent area is highly indulged in the classification

Table 4.1: Number of voxels in each anatomical region of the brain

ROI	Subject 04799	Subject 04820	Subject 04847	Subject 05675	Subject 05680	Subject 05710
CALC	255	408	318	399	334	219
LDLPFC	498	501	440	504	614	478
LFEF	97	128	109	30	71	124
LIPL	299	62	133	263	286	264
LIPS	147	155	235	197	196	226
LIT	366	225	286	233	173	197
LOPER	145	103	169	261	235	101
LPPREC	13	190	153	26	53	110
LSGA	104	55	7	119	95	127
LSPL	171	290	308	265	274	137
LT	340	484	305	476	472	445
LTRIA	190	175	113	136	93	150
RDLPFC	468	374	349	382	453	419
RFEF	81	142	68	45	40	73
RIPL	287	46	90	244	215	213
RIPS	118	65	166	158	237	185
RIT	285	187	277	267	151	209
ROPER	88	140	180	194	255	108
RPPREC	59	124	144	50	72	43
RSGA	71	41	34	54	71	57
RSPL	213	238	252	254	213	119
RT	356	440	286	325	346	357
RTRIA	144	213	57	146	42	118
SMA	154	229	215	77	71	155
Whole brain voxels	4949	5015	4698	5135	5062	4634

task. But it varies from person to person. One specific region has a more significant impact on the determination of sentence polarity for one person but can have minimal impact on another person. When calculating the sum of all active voxels from all subjects in a particular region of interest, there was a clear indication that some specific areas are engaged in the determination of sentence polarity classification tasks. Some parts have little or no contribution to the classification task. The most significant areas resulting from the feature selection voxel list dedicated to determining negative and affirmative polarity classes are CALC, LDLPF, and

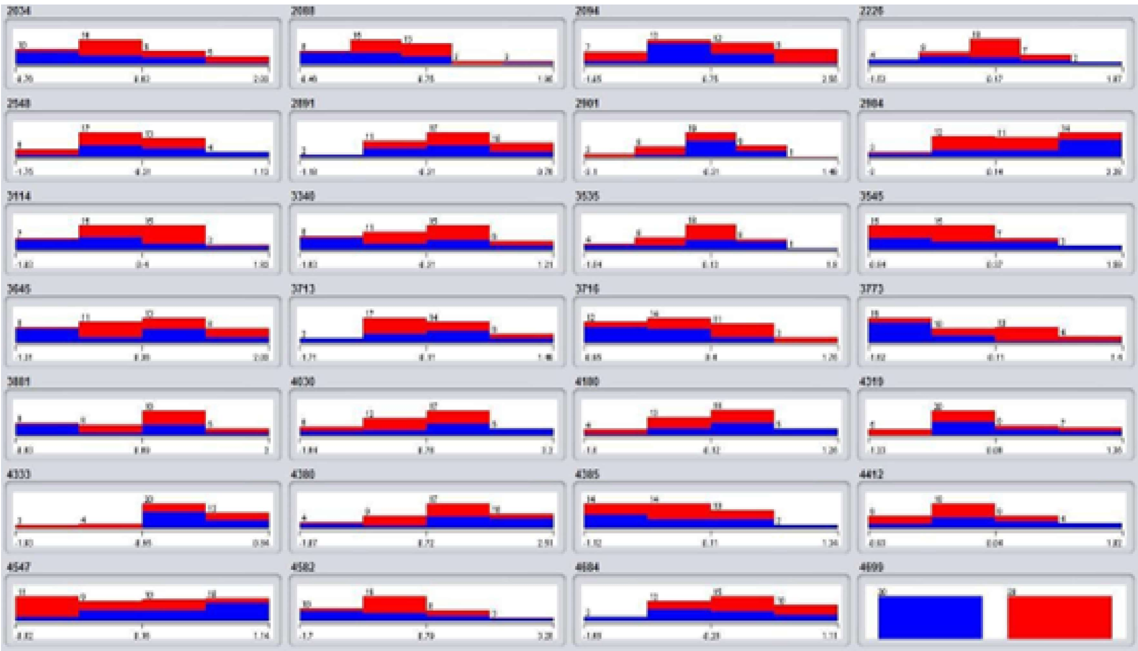

FIGURE 4.5: Statistical distribution and Graphical representations of selected features

RDLPFC. In contrast, the least significant areas are RPPREC, RSGA, and RFEF. The following table 4.2 shows subject wise selected attributes in polarity classification tasks from all regions of the brain. Region-wise number of voxels selected in the determination of sentence classification from all six subjects when summed up category wise becomes -CALC (30), LDLPFC27 LFEF10 LIPL23 LIPS8 LIT19 LOPER9 LPPREC7 LSGA7 LSPL12 LT21 LTRIA7 RDLPFC25 RFEF6 RIPL19 RIPS9 RIT10 ROPER9 RPPREC2 RSGA4 RSPL18 RT24 RTRIA7 SMA9.

The above data clearly shows that the feature selected from CALC, LDLPFC, and RDLPFC area determining features in sentence polarity classification tasks. In contrast, RPPREC, RSGA, and RFEF have minimal impact. All reaming areas of the brain are very little contributing is selecting feature vectors. Their contribution to selecting the central feature voxel is not significant enough. The result obtained is obvious as the maximum numbers of voxel from whole-brain data are when observed in this data set, the major contribution from CALC, LDLPFC, and DLPFC is found.

CFS+RANDOM FOREST The model is trained on the selected feature vector

Table 4.2: Region-wise number of selected features

Area	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
CALC	9	4	4	8	3	2
LDLPFC	3	5	3	5	6	5
LFEF	3	3	1	0	1	2
LIPL	3	1	2	10	7	0
LIPS	2	0	2	3	0	1
LIT	2	1	3	5	3	5
LOPER	0	1	3	0	5	0
LPPREC	0	2	0	1	1	3
LSGA	1	1	2	2	0	1
LSPL	2	3	2	1	3	1
LT	3	3	4	3	4	4
LTRIA	2	1	1	0	1	2
RDLPFC	6	3	3	3	6	4
RFEF	2	2	0	1	0	1
RIPL	9	0	1	4	2	3
RIPS	3	0	0	3	1	2
RIT	4	1	1	2	0	2
ROPER	1	1	2	2	3	0
RPPREC	1	0	0	1	0	0
RSGA	0	1	2	0	1	0
RSPL	2	2	6	4	2	2
RT	4	5	7	2	3	3
RTRIA	0	6	1	0	0	0
SMA	2	3	2	1	0	1

data set and tested on the ten-fold cross-validation approach. The accuracy of the result is good enough and can detect sentence polarity with more than 95% accuracy. Table 4.3 shows the subject wise accuracy result of the sentence polarity task. The average of overall performance analysis in sentence polarity task is represented in fig.4.6, where the performance metrics are as following- TP Rate:0.954167, FP Rate:0.045833, Precision:0.956333, Recall:0.954167, F-Measure:0.954167, MCC:0.9105, ROC Area:0.9885, PRC Area:0.988333, Accuracy:0.954167. The result obtained is quite appreciable as the average classification accuracy of the brain state in the sentence polarity task is more than 95%.

Table 4.3: Result analysis table for all six subjects CFS subset evaluator

Subject Id	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
04799	0.950	0.050	0.955	0.950	0.950	0.905	0.990	0.991	0.95
04820	0.925	0.075	0.926	0.925	0.925	0.851	0.963	0.963	0.925
04847	0.950	0.050	0.950	0.950	0.950	0.900	0.993	0.993	0.95
05675	0.975	0.025	0.976	0.975	0.975	0.951	0.995	0.995	0.975
05680	0.975	0.025	0.976	0.975	0.975	0.951	0.995	0.993	0.975
05710	0.950	0.050	0.955	0.950	0.950	0.905	0.995	0.995	0.95

Table 4.4: Filtered Subset Evaluator

Subject Id	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
04799	0.900	0.100	0.904	0.900	0.900	0.804	0.948	0.948	0.900
04820	0.875	0.125	0.876	0.875	0.875	0.751	0.960	0.959	0.875
04847	0.950	0.050	0.950	0.950	0.950	0.900	0.984	0.984	0.950
05675	0.975	0.025	0.976	0.975	0.975	0.951	0.978	0.979	0.975
05680	0.925	0.075	0.935	0.925	0.925	0.860	0.996	0.995	0.925
05710	0.950	0.050	0.955	0.950	0.950	0.905	0.995	0.995	0.950

Table 4.5: Info-gain

Subject Id	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
04799	0.525	0.475	0.526	0.525	0.522	0.051	0.563	0.587	0.52
04820	0.550	0.450	0.550	0.550	0.550	0.100	0.620	0.610	0.55
04847	0.375	0.625	0.375	0.375	0.375	-0.250	0.381	0.457	0.37
05675	0.500	0.500	0.500	0.500	0.499	0.000	0.526	0.536	0.500
05680	0.475	0.525	0.473	0.475	0.467	-0.052	0.463	0.485	0.475
05710	0.500	0.500	0.500	0.500	0.495	0.000	0.546	0.564	0.500

Table 4.4 shows the result obtained by the filtered subset evaluator, and table 4.5 includes results from info gain feature selection. Comparing the results obtained with different feature selection schemes, our CFS subset evaluator gives the best results among them. The performance of a classifier is judged upon the number of wrongly classified instances, i.e., the error rate or misclassification. The classification model's goal is to determine its ability to correctly classify or predict the class of cases. The confusion matrix explores the error and correctly classified test data in

Table 4.6: Confusion matrix of all subject

Subject id	Predicted Class Affirmative	Negative	Actual Class
04799	18 0	2 20	Affirmative Negative
04820	18 1	2 19	Affirmative Negative
04847	19 1	1 19	Affirmative Negative
05675	19 0	1 20	Affirmative Negative
05680	19 0	1 20	Affirmative Negative
05710	20 2	0 18	Affirmative Negative

terms of true positive, false positive, etc. The confusion matrix displays the number of correct and incorrect predictions in Table 4.6. The column entries represent the predicted class value, and the rows display the actual class values.

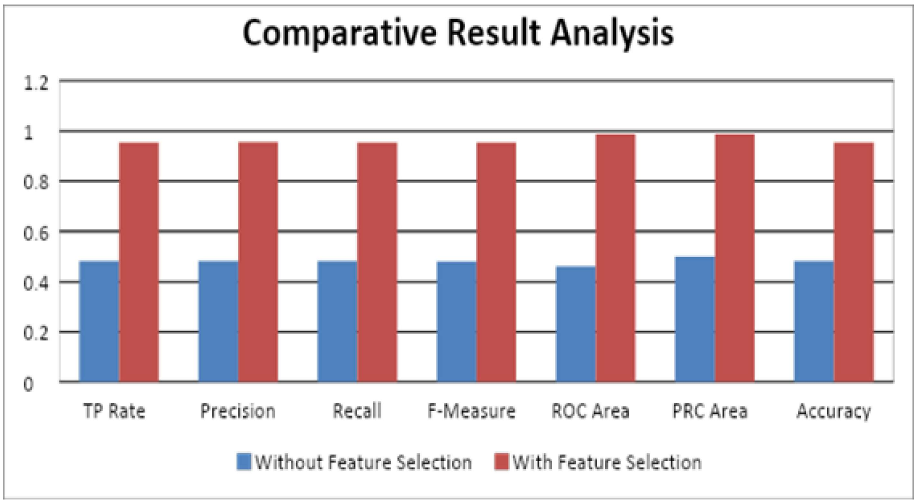

Figure 4.6: Comparative analysis (Overall average of six subjects)

COMPARATIVE RESULT ANALYSIS OF RANDOM FOREST Comparing the result obtained in table 4.7 with table no. 4.3, results indicate that applying feature selection techniques improves the classification result up to 50%. The general opinion that using the feature selection technique causes loss of information, and the result obtained will suffer, does not apply here. We have accomplished much better results with feature selection. When comparing our result with different classification techniques like SVM, Decision tree or Bayesian model we obtained our proposed model gives much better results. The Naive Bayes classifier works on the principle of conditional probability, as given by the Bayes theorem. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression). Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. Definition. Support vector machines (SVMs) are particular linear classifiers which are based on the margin maximisation principle. They perform structural risk minimization, which improves the complexity of the classifier with the aim of achieving excellent generalisation performance.

TABLE 4.7: Result obtained without feature selection from all subjects

Subject Id	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
04799	0.600	0.400	0.601	0.600	0.599	0.201	0.526	0.529	0.60
04820	0.500	0.500	0.500	0.500	0.500	0.000	0.496	0.502	0.50
04847	0.450	0.550	0.448	0.450	0.444	-0.102	0.390	0.463	0.45
05675	0.450	0.550	0.448	0.450	0.444	-0.102	0.575	0.597	0.45
05680	0.450	0.550	0.449	0.450	0.449	-0.101	0.371	0.448	0.45
05710	0.450	0.550	0.450	0.450	0.450	-0.100	0.406	0.463	0.45

k-NN+INFO-GAIN The data for sentence processing is extracted during the result analysis, and then the mean is calculated to minimise the matrix size. The feature set is selected using an info-gain attribute evaluator, which determines the most contributing feature voxels. Choosing the top twenty features from ranked

Table 4.8: Comparative analysis of average performance measures

Classification Method	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Accuracy
Bayesian	0.95	0.05	0.929	0.925	0.925	0.85	0.98	0.98	0.925
k-NN	0.9	0.1	0.906	0.9	0.899	0.85	0.92	0.92	0.900
SVM	0.93	0.06	0.938	0.933	0.933	0.87	0.93	0.90	0.933
Decision Tree	0.60	0.39	0.608	0.604	0.600	0.212	0.61	0.61	0.604
Random Forest	0.95	0.04	0.956	0.954	0.954	0.91	0.98	0.98	0.95

Table 4.9: Result obtained for all Subjects k-NN

Subject id	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Accuracy
04799	0.85	0.15	0.85	0.85	0.85	0.873	0.85
04820	0.9	0.1	0.9	0.9	0.9	0.946	0.90
04847	0.875	0.125	0.884	0.875	0.874	0.889	0.875
05675	0.925	0.075	0.935	0.925	0.925	0.925	0.925
05680	0.9	0.1	0.917	0.9	0.899	0.943	0.90
05710	0.95	0.05	0.955	0.95	0.95	0.99	0.95

feature set classification is done using the k-NN classifier. The result obtained from the classifier is summarised in table 4.9. In subject 05710, the highest classification accuracy is obtained, whereas, in subject 04799, the least classification accuracy is found. To get optimal classification results in accuracy the value of k, i.e., selection of the number of neighbours is tuned properly in the classification.

The confusion matrix describes the performance of the classifier. The confusion matrix for the above classification result is given in table 4.10.

The average of the result obtained for all subjects is projected in fig 4.7. The average result parameters obtained after all subject classification are the following: -TP Rate-0.9, FP Rate- 0.1, Precision-0.906833, Recall-0.9, F-Measure-0.899667, ROC Area-0.927667, and accuracy-0.9.

MLP

Table 4.10: Confusion Matrix

Subject id	Predicted Class Affirmative	Predicted Class Negative	Actual Class
04799	17 3	3 17	Affirmative Negative
04820	18 2	2 18	Affirmative Negative
04847	16 1	4 19	Affirmative Negative
05675	20 3	0 17	Affirmative Negative
05680	20 4	0 16	Affirmative Negative
05710	20 2	0 18	Affirmative Negative

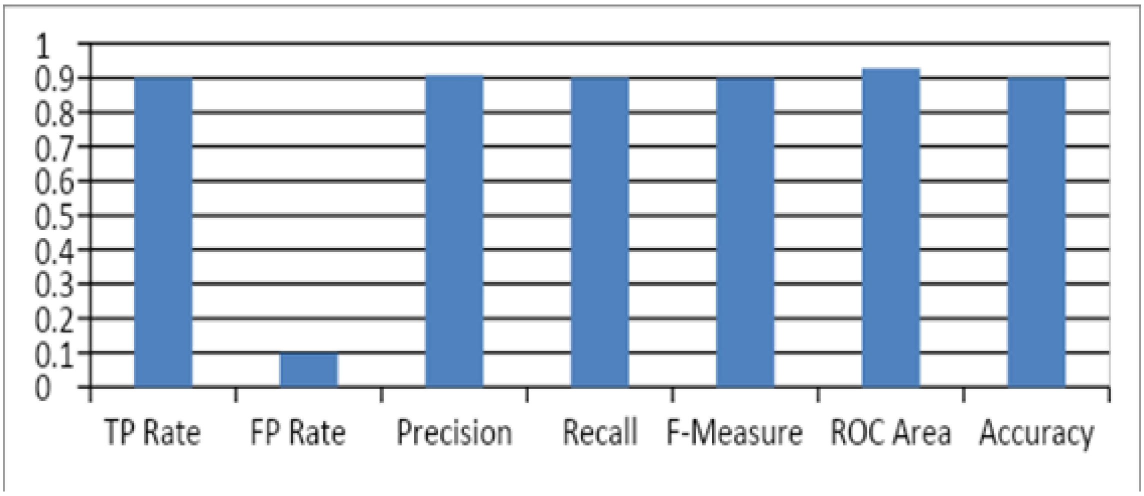

Figure 4.7: Average of the result obtained for all subjects)

The selected feature vectors are fed to the classifier. The classifier is trained to classify the brain state in the processing of affirmative sentences and negative sentences with the assumption that- a) fMRI data has sufficient information to find out the state of the brain corresponding to specific sentence type b) Machine learning algorithm is well efficient to learn the particular temporal pattern to distinguish between the processing of negative vs. affirmative sentences. MLP classification technique is applied to classify the selected feature vectors. The gradient descent method is used for updating weights and bias during the training of the MLP. Bias nodes are added to increase the model's flexibility to fit the data by allowing the activation function to be shifted to the left or right. Initially, the network is initialised with small random weights, and then gradient descent is used to tune the weights into optimised values. Initially, the bias value is set to 0, and a later gradient descent optimizer is used to tune the bias. After the training, the model can detect the brain state corresponding to affirmative or negative sentences. The performance of a classifier is judged upon the number of wrongly classified instances, i.e., the error rate or misclassification. The classification model's goal is to determine its ability to classify or predict the class of instances correctly. The confusion matrix explores the error and correctly classified test data in terms of true positive, false positive, etc. The confusion matrix displays the number of correct and incorrect predictions in Table 4.11. The column entries represent the predicted class value, and the rows display the actual class values. The model is trained on the selected feature vector data set and tested on the ten-fold cross-validation approach. The accuracy of the result is good enough and can detect sentence polarity with more than 93% accuracy. Table 4.12 shows the subject-wise accuracy result of the sentence polarity task. The average of overall performance analysis in the sentence polarity task is represented in fig. 4.8, where the performance metric are as following- TP Rate: 0.933333, FP Rate: 0.066667, Precision: 0.940167, Recall: 0.933333, F-Measure: 0.933, , ROC Area: 0.980167, Accuracy: 0.933333.

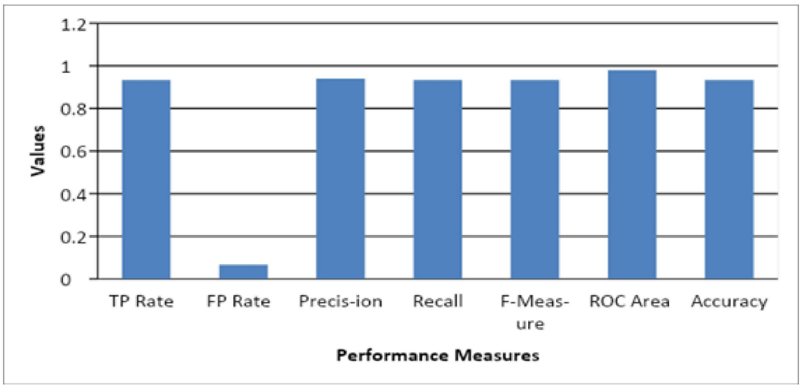

FIGURE 4.8: Average of results for all subjects

Table 4.11: Confusion Matrix

Subject id	Predicted Class Affirmative	Predicted Class Negative	Actual Class
04799	16 0	4 20	Affirmative Negative
04820	19 1	1 19	Affirmative Negative
04847	18 0	2 20	Affirmative Negative
05675	19 0	1 20	Affirmative Negative
05680	20 4	0 16	Affirmative Negative
05710	18 1	2 19	Affirmative Negative

Table 4.12: The classification result

Subject	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Accuracy
04799	.90	0.1	0.917	0.9	0.899	0.993	0.900
04820	.95	.05	0.95	0.95	0.95	0.995	0.950
04847	.95	0.05	0.955	0.95	0.95	0.963	0.950
05675	.975	.025	0.976	0.975	0.975	0.995	0.975
05680	.90	.1	0.917	0.9	0.899	0.95	0.900
05710	.925	.075	0.926	0.925	0.925	0.985	0.925

The result obtained is quite appreciable as the average classification accuracy of the brain state in the sentence polarity task is more than 93%.

ANALYSIS OF RESULTS

We have calculated, and compared Kappa statistics and root means squared error in the determination of the efficiency of a particular classification method. Comparing the above statistics, the Random Forest was found to be a more efficient classification method.

To the determination of the efficiency of the feature selection method, we performed an ANOVA test using Tukey's HSD (honestly significant difference) among different

Table 4.13: Error Analysis

Methods	Kappa statistic	Root means squared error	Root relative squared error
Random Forest	0.9	0.2208	44.1584 %
k-NN	0.75	0.3451	69.0257 %
MLP	0.85	0.2259	45.1868 %

feature selection results, and we obtained the following results-

Table 4.14: Statistical Analysis of Results

Result Analysis	1	2	3	Total
N	6	6	6	18
X	5.725	2.915	5.575	14.215
Mean	0.9542	0.4858	0.9292	0.79
X2	5.4644	1.4354	5.1869	12.0867
Std.Dev.	0.0188	0.062	0.0368	0.225

The f-ratio is 225.03677. The p-value is < 0.00001. The result is significant at $p < .05$

Tukey's HSD (honestly significant difference) procedure facilitates pairwise comparisons within the ANOVA data. The F statistic (above) tells us whether there is an overall difference between the sample means. Tukey's HSD test allows us to determine which of the various pairs of means - if any of them - there is a significant difference. A bold value for Q (below) indicates a considerable result. As the p-value obtained is smaller than 0.05, which indicates the efficiency of T2 (CFS) feature selection over info-gain(T1) and the Filtered method(T3).

SVM-RFE+ROTATION FOREST

With SVM-RFE feature selection and Rotation Forest Classification we obtained 100% accuracy in classification. Initially, we tested our results on default parameters only and that gave 100% accuracy. We tested the data again with SVM-RFE and Rotation Forest with the detailed analysis, and we achieved the following results-

Table 4.15: Pairwise comparisons and p-value

Pairwise Comparisons		HSD.05 =0.0645 HSD.01 =0.0849	Q.05 =3.6734 Q.01 = 4.8359
T1:T2	M2 =0.95 M1 = 0.49	0.47	Q=26.67 (p =.00000)
T2:T3	M2 =0.95 M3 = 0.93	0.03	Q = 1.42 (p = .58418)
T1:T3	M1 =0.49 M3 = 0.93	0.44	Q=25.25 (p =.00000)

Table 4.16: Result of Rotation Forest with Varying feature set

No. of features	10	20	50	100	200	400	1000	2000	5000
Accuracy	100	100	100	98.6	97.5	93.2	85.3	73.2	52.4

1. Based on no. of features fed to the classifier

With the increasing numbers of feature set the efficiency gradually goes down as shown in Table 4.16.

2. Based on test-method-With k-fold cross-validation with 100 features, we got 98.6% accuracy. But with a split of 75% data for training and 25% for testing, we got 97.2% accuracy.

3. However, when we tested the combination of these methods on other data like Noun Classification and Reading Harry Potter Chapter, we could not achieve more than 70% accuracy. This shows the SVM-RFE and Rotation Forest are giving significant results with the Star-Plus data set only.

COMPARATIVE STUDY

Similar studies on the same data set on sentence polarity judgement task is found in literature.

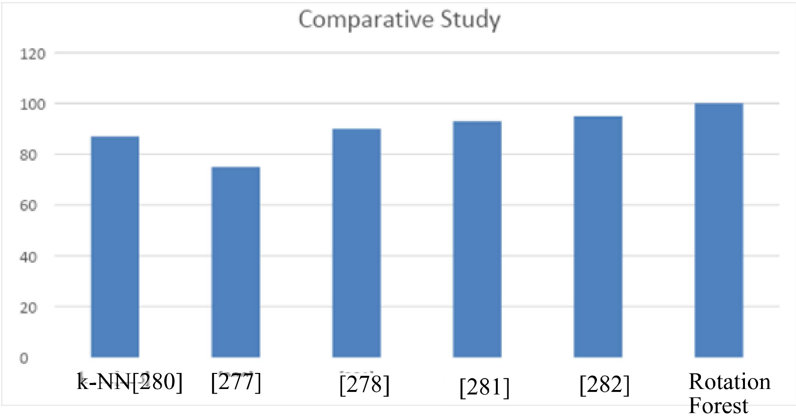

FIGURE 4.9: Comparative result analysis on Star-Plus Dataset

Behroozi et al. [277] have classified the voxel from the RDLPFC area and achieved an accuracy of almost 75% for all the subjects on the same data set. They have used F-score based feature selection and SVM classification techniques in the classification of selected attributes. In [278], authors have studied the neural activity of affirmative and negative sentences and using NeuCube[279] on the same data set. We have compared their results with the obtained results in our approach, and a comparative graph of classification accuracy has been plotted in Fig. 4.9

4.4 CONCLUSIONS

In this chapter we analysed the sentence polarity task on the Star Plus data set. We have figured the sentence polarity using a combination of different sets of classification and feature selection techniques: k-NN+ info gain, MLP+ Entropy based, Random Forest+ CFS and finally Rotation Forest + SVM-RFE. We obtained 100% accuracy with SVM-RFE+ Rotation Forest for some set of attributes but it decreases gradually with the increased number of features fed to the classifier. At last, we compared our results with existing methods in literature on the same data set.