IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 19, 2017, accepted April 12, 2017, date of publication April 27, 2017, date of current version July 31, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2698068

Prediction of Human Brain Activity Using Likelihood Ratio Based Score Fusion

RAHEEL ZAFAR[1], SARAT C. DASS[2], AAMIR SAEED MALIK[1], NIDAL KAMEL[1], M. JAVVAD UR REHMAN[2], RANA FAYYAZ AHMAD[1], JAFRI MALIN ABDULLAH[3], AND FARUQUE REZA[3]

[1]Department of Electrical and Electronic Engineering, Universiti Teknologi Petronas, Perak 32610, Malaysia
[2]Department of Fundamental and Applied Sciences, Universiti Teknologi Petronas, Perak 32610, Malaysia
[3]Center for Neuroscience Services and Research, Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia

Corresponding author: Aamir Saeed Malik (aamir_saeed@utp.edu.my)

This work was supported in part by the Ministry of Education Malaysia through the Higher Institution Center of Excellence for CISIR under Grant 0153CA-005 and in part by the ERGS under Grant 0153AB-i12.

ABSTRACT Human brain has a complex structure with the billions of neurons, so it is a difficult and challenging task to predict the behavior of human brain. Different methods and classifiers are used to measure and classify the brain activities with higher accuracy and reliability. In this paper, instead of using mostly used classifier (support vector machine), prediction of the brain activity is done by estimating the match score densities. This method is based on likelihood ratio test which helps in finding the optimal combination of match scores. The distributions of match scores are modeled for different classes based on density score fusion in which the densities of different classes are estimated from the training data set and match scores are found by fusing the estimated densities with the testing data. The fusion is done with the data extracted from distributed activation patterns using multivariate pattern analysis (MVPA) against a visual task. MVPA is an intense strategy which helps in better understanding of the human brain. The match score-based technique is used in different biometric systems but never been used for the prediction of brain activity. In order to test the performance of proposed method, prediction accuracy is compared with the support vector machine using two data sets of different modalities, one is electroencephalography (EEG) and the other is functional magnetic resonance imaging (fMRI). The results show that the proposed method predicts the novel data with improved accuracy of 66.1% and 69.3% compared with support vector machine which have 64.15% and 65.7% for fMRI and EEG data sets, respectively.

INDEX TERMS fMRI, EEG, likelihood ratio test, SVM, features, classification.

I. INTRODUCTION

In recent years, machine learning and brain imaging provide a platform to study the behavior of brain for different cognitive tasks and diseases [1], [2]. It also helps in the development of brain computer interface or brain machine interface (BCI / BMI) applications [3]. Different modalities like functional magnetic resonance imaging (fMRI), electroencephalography (EEG), positron emission tomography (PET), magnetoencephalography (MEG) and different machine learning techniques like support vector machine (SVM), Gaussian naive Bayes (GNB) and linear discriminant analysis (LDA) are used for the analysis of brain data. All the machine learning techniques differentiate the brain activity patterns against different tasks including cognitive, auditory and visual task or against the particular disease.

The brain is the most complex biological structure on earth which contains billions of neurons, therefore it is always a challenging task to measure and analyze brain activities. Moreover, the measured difference between brain activities with existing modalities is very small and the signal which is of interest is always weak. Prediction of brain is defined in a broad sense by Bubic et al [4]. ‘‘Predictive processing refers to any type of processing which incorporates or generates not just information about the past or the present, but also future states of the body’’. This prediction of brain activity helps in different aspects such as cognitive control, perception and other cognitive processes. Powerful statistical analysis and advanced modality is required to predict the brain behavior in a better way. Modalities are improving rapidly, like in fMRI 7 Tesla machine and in EEG 256 channels system are available to measure brain activities. On the other side, researchers are trying to apply statistical methods in a more significant way for the better prediction of human brain activity. The main steps during the prediction of brain activity include collection of data, artifact removal, and selection of features for classification or prediction. Although, careful data collection and pre-processing have an important contribution in better prediction, but improvements are mostly dependent on the statistical analysis of the data which includes feature extraction, feature selection and classification. Different methods of feature extraction, selection and classification are reported in the literature, however in this study the main focus is on classification/prediction techniques. The classification analysis includes pattern classification in which brain patterns are examined to study the behavior of the brain by classifying the brain states against different task conditions [5]–[9]. The purpose of classifier is to predict the different task conditions against which the individual brain data was acquired; hence classification is a way to find the difference between the conditions.

Different studies used different classifiers according to their requirements and data sets and some studies also compared these classifiers based on brain data. Most of the studies used fMRI data as it is the most popular and reliable technique in brain studies [10], [11]. These studies stated that linear multivariate methods such as linear kernel SVM and LDA are better as compared to the most conventional methods like K-nearest-neighbors and GNB [10]. In this study, it is also shown that nonlinear kernel SVM is not better than linear kernel SVM [10], however another study proved that nonlinear kernel SVM can outperform linear kernel SVM [11]. In short, different classifiers are used in existing brain studies to predict the behavior of brain. The purpose of each study is to predict the behavior of brain with better accuracy. The results are mostly dependent on the data because every study used different experiment, conditions and subjects for data collection.

The primary objective of neuroscientists is to achieve maximum prediction accuracy from human brain activity. For this purpose, different statistical techniques are used along with latest machines for data collection. Recently different new techniques are developed to predict the human brain activity due to its various applications including BCI and diagnosis of diseases especially for epileptic patients [12], [13]. Sparse Bayesian is one of them in which the model parameters can be estimated efficiently under the Bayesian framework [14]–[16]. In neuroscience, during prediction the common way is to test the novel data on a trained classifier. For this purpose, different classifiers are used with different techniques e.g., in SVM the weights are defined which represent the hyperplane. This hyperplane separates the classes as best as possible. Logistic regression (LR) and Naive Bayes (NB) do the predictions using probability. Linear discriminant analysis (LDA) is also a probabilistic approach; however LDA and NB classifiers assume Gaussian within class distributions.

According to literature the most common, widely used and best classifier for brain studies is SVM as it is used in most of fMRI and EEG studies. SVM classifiers are very famous for their generalization ability; therefore it has been used in most of the studies for the analysis of brain data [6], [7], [17], [18].

In the current study, instead of using above discussed classifiers, a different prediction technique is introduced to predict the behavior of the brain. This technique is known as likelihood ratio based score fusion (LRBSF) which is based on density based score fusion and it can directly attain the optimal performance. In this technique, densities of different classes are found using kernel density estimation (KDE) from the training data set only. The match scores (novel testing data) are then fused with these densities of different classes. Finally, the prediction about the class is done using likelihood ratio test (LRT). There are studies which are based on LRT especially in biometric systems [19], [20]. In brain studies, LRT is used for a frequency recognition method to improve the efficiency of SSVEP based BCI [21]. In a fMRI study, generalized LRT is used to find the phase information in the signal [22] and is also used for data analysis to explain the colored noise [23]. This is also used to determine the activations within the voxels [24] and in separation of covariances [25]. However, we have used the combination of KDE and LRT to predict the behavior of brain which is a novel implementation for EEG and fMRI data sets.

This study presents a prediction technique which has been used in different biometric systems and has achieved high recognition rates (accuracy), especially in face matching, finger and speech matching. However, it has never been used in prediction of brain activities. This study is the first attempt to examine the behavior of brain by using KDE and LRT. Moreover, this technique is tested on collected data of two different modalities i.e. EEG and fMRI. To check the performance of this technique, the results are compared with the results found using best existing technique, i.e. SVM. The proposed technique has produced good results compared to SVM for both data sets i.e. EEG and fMRI.

II. LIKELIHOOD RATIO BASED SCORE FUSION

A. LIKELIHOOD RATIO TEST (LRT)

The match scores for K different given matchers are represented by $X = [X_1, X_2, X_3, \cdots, X_k]$. The match score of the kth matcher is represented by the random variable X_k where $k = 1, 2, \cdots, K$. Let the two classes are denoted by M_0 and M_1 where M$_1$ is true positive (genuine) class and M_0 is true negative class. $f_0(x)$ and $f_1(x)$ are the conditional joint densities of the k match scores given the first and second classes respectively, where $x = [x_1, x_2, x_3, \cdots, x_k]$. Suppose that the target is to assign the observed match score vector X to one of two classes i.e. M_0 or M_1. H_0 is the null hypothesis which should be rejected and H_1 is the alternative hypothesis. According to Neyman-Pearson theorem, [26] for testing a hypothesis $H_0 : f = f_0$ against $H_1 : f = f_1$, the likelihood ratio test which rejects H_0 in favor of H_1 has the form

$$\Psi(x) = \frac{f_0(x)}{f_1(x)} \leq \eta \tag{1}$$

$$P(\Psi(x) \leq \eta) = \gamma \tag{2}$$

If the equations (1) and (2) are satisfied for threshold η, it is the most powerful test among all level γ test where γ is false accept rate or type II error. According to the Neyman-Pearson theorem, the ideal test for deciding about the score vector X (testing data set) is the likelihood ratio test (equation 1) which tells that whether the given test vector is from M_0 or M_1 class. For certain threshold η, genuine accept rate (GAR) or accuracy can be maximized using LRT. According to the Neyman-Pearson theorem [26], it is not possible that there is any other decision rule which can maximize the GAR or accuracy. This rule of the likelihood ratio test is dependent upon the underlying densities and guaranteed only when they are known. The optimal test for assigning a score vector x to M_1 and M_0 class is the likelihood ratio test given by $f_1(x)/f_0(x)$, where $f_1(x)$ and $f_0(x)$ are the densities which are estimated from the training data set of M_1 and M_0 match scores. Although Gaussian Naive Bayes is used in brain studies [18] but sometimes the Gaussian density is not appropriate for brain prediction of brain so here the normal distribution is extended to density estimation and the likelihood ratio test is dependent upon the accuracy of these density estimates.

B. KERNEL DENSITY ESTIMATION

Kernel Density Estimation is a non-parametric way to estimate the probability density function (pdf) of a random variable. Without any presumptive distributional properties, it is used for the estimation of distribution on a given set of data samples [27]. The advantage of non-parametric estimation is that they don't have fixed structure and depend upon all the data points for the estimation. KDE is widely used in many fields and applications; e.g. in computer vision it is used to identify the target objects [28], [29]. It also has different application in transportation [30], [31]. Tabibiazar and Basir [30] collected the car data and extracted the congestion spot in road network by using KDE. KDE is also used in the sea traffic to detect anomalies by Laxhammar et al. [31]. It is also used in particle smoothing [32] of fMRI data, but we have used KDE with LRT to predict the M_1 and M_0 class by giving the novel data set of brain. KDE is used to approximate the pdfs of the training data for both classes in such a way that the tested data can be estimated with maximum likelihood. KDE is an appropriate tool because with any statistical model the patterns are not consistent in any application. So it is difficult to know in advance, that which model describes the distribution in a more appropriate way.

Let the set of data samples is $[x_0, x_1, \cdots, x_n]$. To generate a discrete probability mass function, a histogram can be used. A pdf by the estimator can be explained in the following way [33]

$$\hat{f}_{KDE}(x) = \frac{1}{n}\sum_{i=1}^{n} K_a(x - x_i) = \frac{1}{na}\sum_{i=1}^{n} K\left(\frac{x - x_i}{a}\right) \quad (3)$$

Where $\mathrm{x_i}$ represents a data sample and i has an integer value maximum up to number of samples. K is the kernel function. $K_a(x)$ is a symmetric pdf such as Gaussian and n is the sample size. The width of the kernel function is determined by the smoothing parameter a. The shape of the distribution to be estimated is approximated by the sum. In this paper we have used normal kernel smoother and standard Gaussian kernel function $N(0, 1)$ for the implementation as follows.

$$K\left(\frac{x - x_i}{a}\right) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(\frac{x - x_i}{a}\right)^2} \quad (4)$$

C. FUSION OF MATCH SCORES BASED ON LIKELIHOOD RATIO TEST

To find the prediction accuracy with test data, match scores vector is fused with quality vector estimated from training data set of both classes. The quality based likelihood ratio is found using the equation 1 which should be greater than 1 for M_1 and less than 1 for M_0 class. In this technique, density based scores are fused and it needs explicit estimation of M_1 and M_0 match score densities. The advantage of this approach is that it can directly attain ideal performance for any desired point and is based on the estimation accuracy of the scores. The KDE is used to estimate the densities of the classes.

D. FUSION OF MATCH SCORES AND THEIR PERFORMANCE

To define the fusion of match scores based on likelihood ratio, a vector of match scores and estimated densities are required to compute the likelihood ratio fusion. Let K is match scores vector where $x = [x_1, x_2, x_3, \cdots, x_k]$ and $\hat{f}_1(x)$ and $\hat{f}_0(x)$ are estimated densities of true positive class and true negative class respectively. The x should be assigned to the M_1 class if $LR(x) \geq \eta$, where $LR(x) = \hat{f}_1(x)/\hat{f}_0(x)and\eta$ is the decision threshold which is found based on the overall accuracy.

The quality and performance of the system depends upon the accuracy of the matchers and it is difficult to classify a poor quality sample as either a M_1 or M_0 sample. The likelihood ratio of these types of samples are closer to 1 while the good samples have greater than 1 for M_1 class and less than *1* for M_0 class. The likelihood ratios are weighted by the respective sample quality during the estimation of the joint density of match score and the associated quality.

Let $Y = [Y_1, Y_2, Y_3, \cdots Y_k]$ is a quality vector where the quality of the match score for $\mathrm{k^{th}}$ matcher is denoted byY_k. $\hat{f}_1(x, y)$ and $\hat{f}_0(x, y)$ are the joint densities of match score vector with K-dimension and quality vector of the same dimension. These quality vectors are estimated from M_1 and M_0 classes. The quality based likelihood ratio $Q(x, y)$ can be explained in the following way

$$Q(x, y) = \frac{\hat{f}_1(x, y)}{\hat{f}_0(x, y)} \quad (5)$$

The decision of assigning $Q(x, y)$ to M_1 class is done when $Q(x, y) \geq \eta$ which is same as we discussed earlier for $LR(x) \geq \eta$. The joint density estimation of (X, Y) has $2K$ variables which are not reliable with less training data. Independence of K matchers are assumed to avoid the curse of dimensionality so the above equation is written as

$$\hat{f}(x,y)=\prod_{k=1}^{K}\hat{f}_k(x_k,y_k)=\prod_{k=1}^{K}\frac{\hat{f}_{1,k}}{\hat{f}_{0,k}} \tag{6}$$

III. MATERIAL AND METHODS

A. SUBJECTS

Data of eight subjects were used for the final analysis after excluding criteria. One of the subjects had excessive head movement in the scanner and two subjects identified the categories with very low accuracy. All students had submitted the written consent form before the start of the experiment. All subjects were from 21 to 26 in age and data of 4 male and 4 female was included in the final analysis. The study protocol was approved by the local Hospital Universiti Sains Malaysia (HUSM) ethics committee.

B. STIMULI AND TASK

A total of 260 grayscale photographs were presented in three different sessions. All images were taken from internet, freely available and had already been used in a previous study [34]. Every image was of size 500 × 500 pixels with at 4 × 4 pixels fixation spot in the middle of every image.

Both EEG and fMRI data was collected in three short separate sessions. Duration of each session was around 15 mins and approximately 100 images were presented twice in each session. This data was collected simultaneously so both EEG and fMRI have same parameters. After removing images due to scanner and jittering effect, a total of 260 images were left out for final analysis. These images were further divided into five categories, human, animal, building, natural scenes and fruits.

C. FUNCTIONAL MRI / MRI DATA ACQUISITION

The study was conducted at Hospital Universiti Sains Malaysia (Department of Radiology). The EEG recordings and fMRI data acquisition were taken place at HUSM. In fMRI, a trigger code was sent from the scanner to start the visual presentation.

The inter stimulus interval (ISI) for data collection was 4s; with 1s image and 3s rest period for BOLD response. Stimulus was presented for 1s with 200ms on and off. Data was acquired on a Philips Acieva 3T scanner using a gradient echo EPI pulse sequence $\mathbf{TR = 2000ms}$, $\mathbf{TE = 30ms}$ and with voxel size of 3 × 3 × 3 mm having 35 slices. The anatomical data was recorded for 5 mins with voxel size of 1.1 × 1.1 × 1.2 mm, $\mathbf{TR = 7.5ms}$, $\mathbf{TE = 3.5ms}$. To overcome the jittering effect an additional delay was given after every 8th image and some blank images were presented at the start of the experiment which was ignored in the analysis to avoid the scanner effect. Jittering is to vary the time of TR relative to the stimulus presentation and is done to get a design with very high statistical power and efficiency. TR represents repetition time, which is the amount of time that passes between consecutive excitation pulses.

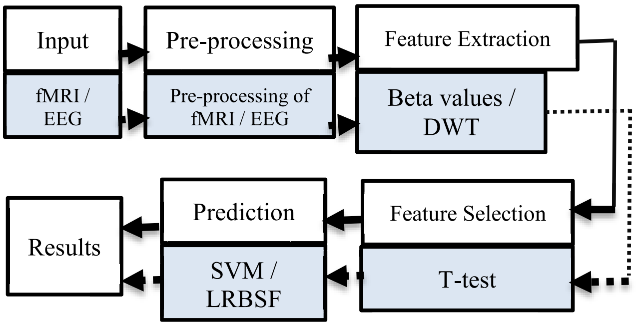


FIGURE 1. Block diagram of proposed method for brain prediction.

D. EEG DATA ACQUISITION

Continuous EEG data was recorded with a 128 channel Electrical Geodesics Incorporated (EGI) system with frequency of 250 Hz. The electrodes were placed according to an extended international 10–20 system with reference electrode positioned at FCz. Additional electrode was placed beneath the participant's left chest to monitor electrocardiograms (ECG). The recorded EEG signal was filtered and transmitted to a mac system. Net Station 5 software was used to record, save and display the online recordings of EEG data.

E. PROPOSED METHOD

The proposed method follows the normal prediction steps; preprocessing, feature extraction, feature selection and prediction. The block diagram of proposed method is shown in Figure 1 which explains the complete prediction process. The same steps are followed for both data sets and the performance of the proposed technique i.e. LRBSF is evaluated on two collected data sets, one is fMRI data and the other is EEG. The performance of this technique is compared with the results found using SVM. The same features and cross validation (Monte Carlo CV with 100 repetitions) is used for the comparison. The only difference in comparison is at last step where LRBSF is used as proposed technique instead of SVM for final results. Moreover, for each classifier the accuracy is found for different number of features in both data sets.

F. PRE-PROCESSING AND FEATURE EXTRACTION OF fMRI DATA

The complete processing i.e. preprocessing and statistical analysis of the fMRI data was done using SPM 8 (Wellcome Department of Imaging Neuroscience, University College London, UK). The preprocessing steps done before statistical analysis are as follows. Slice time correction, which corrects the differences in image acquisition time between slices. Realignment, which do motion correction using six parameters. These parameters are later modeled as nuisance regressors during statistical analysis [35]. Co-registration, in which anatomical image is co-register with functional image for better results. Normalization, in which every brain is normalized using MNI template (Montreal Neurological

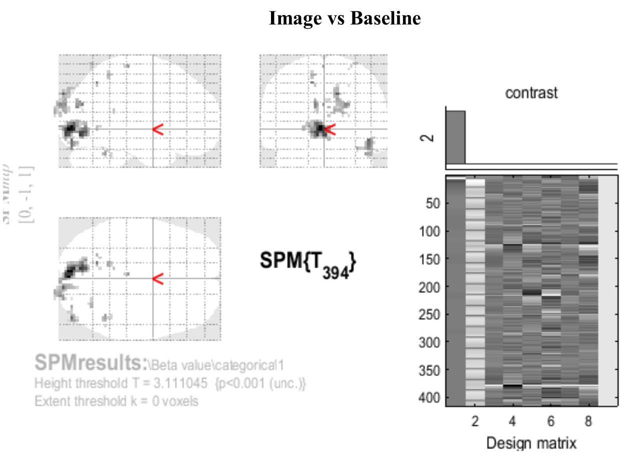


FIGURE 2. Analysis and design matrix of one of the conditions.

Institute). Spatial smoothing, in which every image is smoothed using a Gaussian kernel with full-width at half maximum (FWHM) of 8mm. The same high pass filter is applied to the design matrix during the trial-by-trial estimation.

The BOLD signal change obtained from fMRI recording is fitted with an optimized model of two gamma function (Canonical HRF) which gives estimation value β for each trial and each voxel. These beta values are used as features which are further reduced using the two-sample t-test. To predict fMRI signal, the time series of stimulus onsets is convolved with this canonical HRF because gamma HRF does not allow for a post peak undershoot. All beta values are placed in a row vector and separate design matrix is designed for every image as did by Mumford et al [36]. The dimension of estimated value for each image is $63 \times 53 \times 46$ which is arranged in a row vector for further analysis. The design matrix is shown in a better way along with the brain response against the task in Figure 2. This figure is only for one condition and session for the first subject, however the rest subjects and sessions have activation in the similar region of brain. The final matrix has estimated values (features) in the columns against each image in the row. The same pattern is also followed for EEG data analysis where features (DWT coefficients) are placed in the columns against each image in the row.

G. PRE-PROCESSING AND FEATURE EXTRACTION OF EEG DATA

The pre-processing was done in Brain Electrical Source Analysis (BESA) software where it was filtered from 0.5 to 30 Hz frequencies as these were the most relevant frequencies during a visual task [37]. Eye blinks artifacts were corrected using adaptive artifact correction method present in BESA. Further artifacts were corrected by selecting the unwanted signal manually. In BESA, we can remove any type of artifacts by selecting default block epoch and remove the unwanted patterns. After removing artifacts, the file was exported to MAT LAB for further analysis.

TABLE 1. Number of decomposed features from a single image using DWT.

Wavelet	Db4
D1	128×128
D2	128×67
D3	128×37
D4	128×22
D5	128×14
A5	128×14
Total	128×282

Normally, EEG signals can better be explained in the frequency domain. The most common frequencies are delta (1-4 Hz), theta ($4-8Hz$), alpha ($8-13Hz$), beta ($13-30Hz$) and gamma ($> 30Hz$). The common approaches of feature extraction in frequency domain are Fourier transforms (FT), short time Fourier transforms (STFT) and wavelet transforms (WT). WT can better handle the non-stationary signals as it has the information of both time and frequency domain. The EEG signal can better be explained using WT instead of FT [38], [39]. In DWT, the signal is analyzed at different frequency bands which decompose the signal into a detail and an approximation component. In this study we have used 128 channel system with frequency $250Hz$ so we have 282 DWT features for every image which is mention in Table 1.

H. REDUCTION OF DATA SETS / FEATURE SELECTION

After preprocessing and extraction of features, the main step is the selection of significant features. Since both data sets have high dimension, therefore the significant features were found from both data sets. During feature selection, the n most significant features between the categories were extracted using t-test. The features with p value of $p < 0.05$ were considered as significant. We ranked the features with lowest p values and different subsets were made from 50 to 2, 000 features. The features with $p > 0.05$ are excluded from the analysis because they do not have much impact during analysis. Ranking of features mean the most important features [40] which are selected with t-test.

I. STATISTICAL TESTING

A common technique to check the significance of the study is to use cross validation. During cross validation, part of the available data is used to train the model and some part of the data is used for the testing of the model. The aim of cross validation is to make the classifier unbiased. If cross validation is biased means to use the knowledge of training set in the test or in other words the test set data is not independent with respect to the training data, then the purpose of cross validation do not achieve. In that case, the cross validation error does not reflect the true generalization error

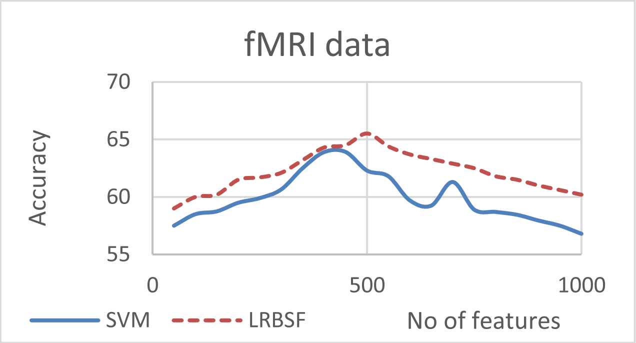


FIGURE 3. **Predictive accuracy with different number of features for fMRI data.**

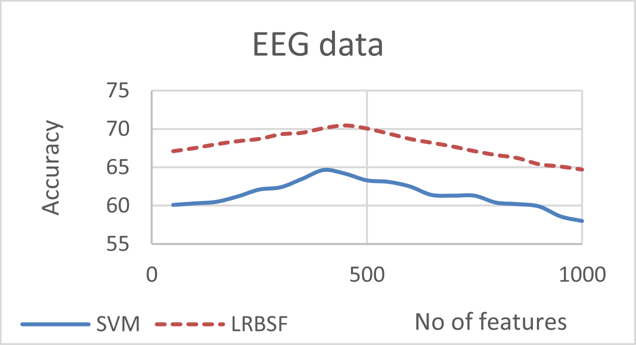


FIGURE 4. **Predictive accuracy with different number of features for EEG data.**

of the model. In this study, Monte-Carlo cross validation procedure was applied to check the performance of SVM and the proposed method. The process was repeated 100 times and every time the entire data was divided randomly into 90% training set and 10% testing set. Since we have 260 images (samples) and 5 classes, so for each one to one combination, we had approximately 104 images/samples. In all cases 10 images/samples were used for testing while rest were used for the purpose of training.

Decoding accuracy can be achieved by classifying the correct test vectors or using confusion matrix to assess the performance of the classifiers such as SVM or LRBSF. In confusion matrix, the frequencies of all classified experimental conditions are presented including the details of misclassification. Misclassification shows which conditions are most distinct and which are more similar. In this study, the data is shuffled 100 times before every classification to get an overall accuracy (randomness) of the system and average is taken for the final results. This shuffling avoids the distribution of accuracy that would happen by chance.

IV. RESULTS

The data was collected from both modalities for five different categories. The prediction was done with SVM and LRBSF on novel data set. Initially, the significant data (after the selection of significant features) were given to the SVM with linear kernel to predict the testing data set. All conditions and subjects were evaluated individually. After that the same data were given to the proposed method to find the prediction accuracy. Different number of features were used for the analysis as discussed above and the features with best accuracy were taken for the final results. The best results were with 400-600 features in both data sets which are shown in Figure 3 and 4 for fMRI and EEG data respectively. For one to one comparison we have used the formula $m(m-1)/2$ where m is the number of categories; for $m = 5$, there were 10 different conditions for every subject. In each comparison; there were two classes with 104 images approximately. Data was divided into 90% training and 10% testing in all cases, while the training-testing partitioning was repeated 100 times randomly by using Monte Carlo cross validation. For each repetition, the densities were estimated from the training data set only.

Figure 5 and 6 explain the average accuracy of eight subjects across different conditions. Figure 5 shows the average accuracy of both methods for fMRI data set, while Figure 6 shows the average accuracy of both methods for EEG data set. The red and blue lines are showing the average accuracy of all subjects for each condition separately with proposed system i.e. likelihood ratio based score fusion (LRBSF) and LIBSVM respectively. The filled and non-filled red marks show the maximum and minimum accuracy of any individual subject respectively against the given condition with LRBSF. Similarly, the blue filled and non-filled marks shows the same thing with SVM.

Distribution of M_1 and M_0 match score is shown in Figure 7. These results are for one trial where class A and B are representing the M_1 and M_0 match respectively. We have explained the first condition of first subject with fMRI data. Same distribution is seen in rest of the cases, i.e. other conditions, subjects and for EEG data set. In Figure 7, it can be seen that for M_1 and M_0 classes, the bandwidth and amplitude of the curves are different. However, there is an overlapping region between the curves which is the false positive and false negative and can be decided based on some threshold value.

To find the statistical significance of the decoding accuracies between both approaches, paired t-test was used for both data sets. P value of 0.05 was used to control the type I error. Firstly, paired t-test was applied on fMRI accuracy results between SVM and LRBSF for all ten conditions. For one tail, the value of $p(T \leq t)$ was 0.000601 and for two tail it was0.001203. Both were quite significant. Secondly, paired t-test was applied on EEG accuracy results between SVM and LRBSF for all ten conditions. This result was even better compared to the previous one. For one tail, the value of $p(T \leq t)$ was 0.0000246 and for two tail it was 0.00004913.

V. DISCUSSION

In this study, we have introduced a different prediction technique for neuroimaging data which have shown better prediction accuracy than SVM. In this method, prediction of novel data is done based on the estimated densities.

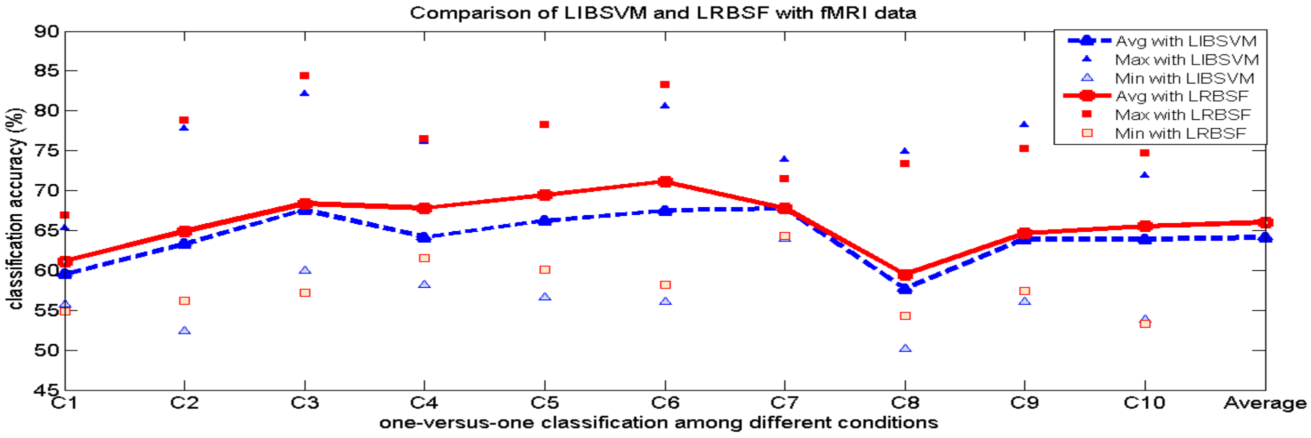


FIGURE 5. Performance of LRBSF and SVM among different conditions on fMRI data.

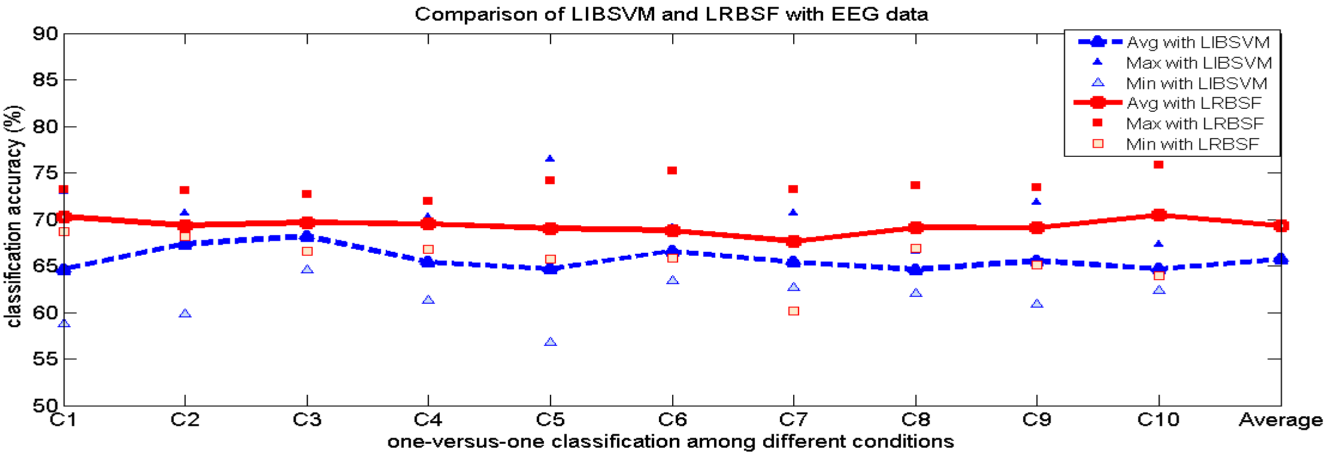


FIGURE 6. Performance of LRBSF and SVM among different conditions on EEG data.

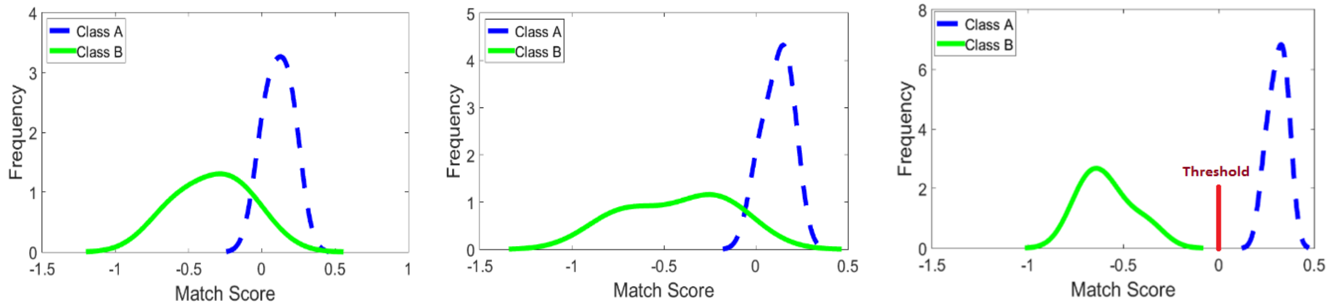
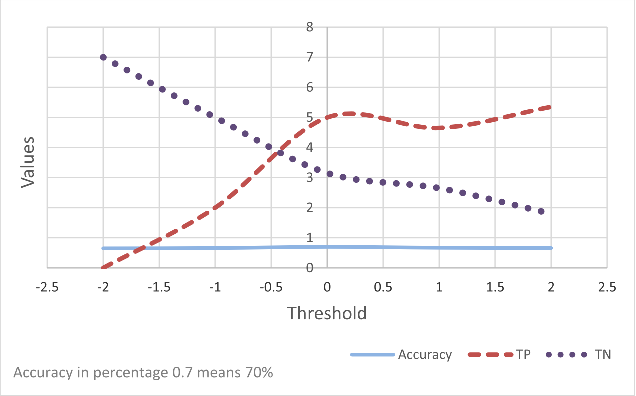


FIGURE 7. Distribution of match scores on fMRI data. From left to right, density estimation cures of two classes for 70%, 80% and 100% (perfect system) accuracy respectively.

Figure. 5 and Figure. 6 express the results of both methods (LRBSF and SVM) and modalities i.e. fMRI and EEG respectively. The comparison shows that LRBSF has better accuracy for both data sets, even for the individual subjects in most of the cases. The five conditions of the experiment are human, animal, building, natural scenes and fruit. The one to one comparison of all conditions with both methods is shown in Table 2. The conditions are in the following order, human vs animal (C1), human vs building (C2), human vs natural scenes (C3), human vs fruit (C4), animal vs building (C5), animal vs natural scenes (C6), animal vs fruit (C7), building vs natural scenes (C8), building vs fruit (C9) and natural scenes vs fruit (C10). From Table 2, it can be observed that the proposed system (LRBSF) have better mean and overall

TABLE 2. **The performance comparison among likelihood ratio based fusion and support vector machine on FMRI and EEG data sets.**

Condition	*fMRI data with SVM (accuracy)*	*fMRI data with LRBSF (accuracy)*	*EEG data with SVM (accuracy)*	*EEG data with LRBSF (accuracy)*
human vs animal	*59.5%*	*61.15%*	*64.6%*	*70.3%*
human vs building	*63.27%*	*64.92%*	*67.4%*	*69.34%*
human vs natural scenes	*67.6%*	*68.37%*	*68.2%*	*69.65%*
human vs fruit	*64.13%*	*67.79%*	*65.4%*	*69.48%*
animal vs building	*66.23%*	*69.42%*	*64.65%*	*69.05%*
animal vs natural scenes	*67.49%*	*71.3%*	*66.63%*	*68.79%*
animal vs fruit	*67.74%*	*67.73%*	*65.42%*	*67.64%*
building vs natural scenes	*57.68%*	*59.47%*	*64.6%*	*69.15%*
building vs fruit	*63.93%*	*64.62%*	*65.53%*	*69.07%*
natural scenes vs fruit	*63.92%*	*65.52%*	*64.65%*	*70.45%*
Average	*64.15%*	***66.10%***	*65.7%*	***69.3%***



Accuracy in percentage 0.7 means 70%

FIGURE 8. **Accuracy, true positive and true negative values for different thresholds.**

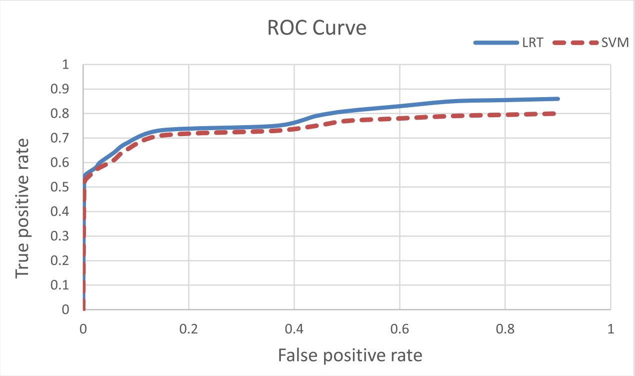


FIGURE 9. **ROC curve of proposed method and SVM.**

results of all subjects and conditions on both data sets.

Figure 7 shows the distribution of both match scores for fMRI data of one session only. The left most is for 70%, the middle one is for 80% and the right most is for 100% accuracy. It can be observed that the behavior is changed with increase in accuracy and the overlapping region between both the curves is decreased. For perfect classification (100% accuracy), there is no overlapping between the curves. These curves are from single run and not the average of overall accuracy.

Different numbers of features were used to find the accuracy using both techniques, i.e. SVM and LRBSF. We have made different subsets of features and in each subset different numbers of features were defined. Minimum 50 and maximum 2, 000 features were used in one subset with an increase of 50 features for next subset compared to the previous one. According to Neyman-Pearson theorem [26], a threshold is needed to decide about the class of a score vector X. This vector X is from the test data set and the threshold has an important role in the decision of likelihood ratio test, because the maximum accuracy about M_1 class is dependent upon this threshold. To find the best threshold of the likelihood ratio test, different values of threshold are used and the best one according to accuracy is chosen for the analysis. After estimating the densities of each test vector from M_1 and M_0 class of the training data set, the next step is to apply likelihood ratio test to find the appropriate class. Accuracy, specificity and sensitivity are found with different threshold values and the best one is chosen with finite sensitivity and specificity values. The changes in all these fields are observed with the change in threshold; accuracy is not much changed but the true positive and true negative values are changed significantly as shown in Figure 8.

The Receiver operating characteristic (ROC) curve is made for different threshold values both for SVM and LRBSF Figure 9, which illustrates the performance of both with varied threshold. This ROC is made based on the average of 100 runs for one condition of fMRI data, while the other conditions and subjects have shown the same pattern with different values of true positive rate and false positive rate.

VI. CONCLUSIONS AND FUTURE WORK

Likelihood ratio based score fusion is a popular and reliable method for biometric systems; however to our knowledge, it has not been used in neuroscience for the prediction of brain activity. In this study, prediction is done based on the features of EEG and fMRI data sets. In this method, the densities are estimated using KDE from the training data of different classes and fused with the testing data to select the best appropriate class of the testing data. The performance of the proposed method is evaluated by comparing the result of proposed method with SVM. SVM is a mature and famous method to classify the brain states; however our proposed method showed better accuracy with both data sets. Additionally, there is no need to carefully select the parameters like C and gamma in SVM.

KDE is a flexible way to estimate the densities and according to Neyman-Pearson theorem, LRT is an optimal test for deciding whether a score vector X corresponds to class A or class B as shown in Figure 7. The combination of both show that it can predict the brain activity with better accuracy and reliability. Finally, it is concluded that the proposed method is a significant and better addition for the prediction of brain activity patterns. Although in this study, LRT has outperformed the SVM but in the future, this study can be improved in the following different ways. In this study, limited number of images were taken (approx. 260) for five classes which can be increased in the future along with the number of classes. Moreover, the LRT can be applied with other density estimation techniques like Gaussian mixture model. Finally, more statistical techniques can be applied on neuroimaging data (like LRBSF in this study) for better prediction.

REFERENCES

[1] G. Yourganov, T. Schmah, N. W. Churchill, M. G. Berman, C. L. Grady, and S. C. Strother, "Pattern classification of fMRI data: Applications for analysis of spatially distributed cortical networks," *Neuroimage*, vol. 96, pp. 117–132, Aug. 2014.

[2] D. Cruse *et al.*, "Bedside detection of awareness in the vegetative state: A cohort study," *Lancet*, vol. 378, no. 9809, pp. 2088–2094, 2012.

[3] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *Int. J. Neural Syst.*, vol. 24, no. 4, pp. 1450013-1–1450013-14, 2014.

[4] A. Bubic, D. Y. von Cramon, and R. I. Schubotz, "Prediction, cognition and the brain," *Frontiers Human Neurosci.*, vol. 4, p. 25, Mar. 2010.

[5] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[6] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature Neurosci.*, vol. 8, pp. 679–685, Apr. 2005.

[7] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nature Rev. Neurosci.*, vol. 7, pp. 523–534, Jul. 2006.

[8] R. Zafar *et al.*, "Decoding of visual information from human brain activity: A review of fMRI and EEG studies," *J. Integr. Neurosci.*, vol. 14, no. 2, pp. 1–14, 2015.

[9] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: Multi-voxel pattern analysis of fMRI data," *Trends Cognit. Sci.*, vol. 10, no. 9, pp. 424–430, 2006.

[10] M. Misaki, Y. Kim, P. A. Bandettini, and N. Kriegeskorte, "Comparison of multivariate classifiers and response normalizations for pattern-information fMRI," *Neuroimage*, vol. 53, no. 1, pp. 103–118, 2010.

[11] T. Schmah, G. Yourganov, R. S. Zemel, G. E. Hinton, S. L. Small, and S. C. Strother, "Comparing classification methods for longitudinal fMRI studies," *Neural Comput.*, vol. 22, no. 11, pp. 2729–2762, 2010.

[12] M. I. Khalid *et al.*, "Epileptic MEG spikes detection using common spatial patterns and linear discriminant analysis," *IEEE Access*, vol. 4, pp. 4629–4634, 2016.

[13] A. Sharmila and P. Geethanjali, "DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers," *IEEE Access*, vol. 4, pp. 7716–7727, 2016.

[14] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, p. 1650032, Jul. 2016.

[15] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse Bayesian classification of EEG for brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2256–2267, Nov. 2016.

[16] Y. Zhang, G. Zhou, J. Jin, Y. Zhang, X. Wang, and A. Cichocki, "Sparse Bayesian multiway canonical correlation analysis for EEG pattern recognition," *Neurocomputing*, vol. 225, pp. 103–110, Feb. 2017.

[17] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) 'brain reading': Detecting and classifying distributed patterns of fMRI activity in human visual cortex," *Neuroimage*, vol. 19, pp. 261–270, Jun. 2003.

[18] T. M. Mitchell *et al.*, "Learning to decode cognitive states from brain images," *Mach. Learn.*, vol. 57, no. 1, pp. 145–175, 2004.

[19] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 342–347, Feb. 2008.

[20] L. Nanni, A. Lumini, and S. Brahnam, "Likelihood ratio based features for a trained biometric score fusion," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 58–63, 2011.

[21] Y. Zhang, L. Dong, R. Zhang, D. Yao, Y. Zhang, and P. Xu,"An efficient frequency recognition method based on likelihood ratio test for SSVEP-based BCI," *Comput. Math. Methods Med.*, vol. 2014, Aug. 2014, Art. no. 908719.

[22] J. Sijbers and A. J. D. Dekker, "Generalized likelihood ratio tests for complex fMRI data: A Simulation study," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 604–611, May 2005.

[23] J. Sijbers, A. J. den Dekker, and R. Bos, "A likelihood ratio test for functional mri data analysis to account for colored noise," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2005, pp. 538–546.

[24] D. B. Rowe and B. R. Logan, "A complex way to compute fMRI activation," *Neuroimage*, vol. 23, no. 3, pp. 1078–1092, 2004.

[25] M. W. Mitchell, M. G. Genton, and M. L. Gumpertz, "A likelihood ratio test for separability of covariances," *J. Multivariate Anal.*, vol. 97, no. 5, pp. 1025–1043, 2006.

[26] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY, USA: Springer, 2005.

[27] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, vol. 26. Boca Raton, FL, USA: CRC Press, 1986.

[28] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.

[29] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *J. Roy. Statist. Soc. Ser. B (Methodological)*, vol. 53, no. 3, pp. 683–690, 1991.

[30] A. Tabibiazar and O. Basir, "Kernel-based optimization for traffic density estimation in its," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2011, pp. 1–5.

[31] R. Laxhammar, G. Falkman, and E. Sviestins, "Anomaly detection in sea traffic—A comparison of the Gaussian mixture model and the kernel density estimator," in *Proc. 12th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2009, pp. 756–763.

[32] L. Murray and A. Storkey, "Particle smoothing in continuous time: A fast approach via density estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1017–1026, Mar. 2011.

[33] Y. Cao, L. Zhang, and D. Sun, "An air traffic prediction model based on kernel density estimation," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 6333–6338.

[34] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, pp. 352–355, Mar. 2008.

[35] R. Zafar, A. S. Malik, N. Kamel, and S. C. Dass, "Importance of realignment parameters in fMRI data analysis," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Oct. 2015, pp. 546–550.

[36] J. A. Mumford, B. O. Turner, F. G. Ashby, and R. A. Poldrack, "Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses," *NeuroImage*, vol. 59, no. 3, pp. 2636–2643, 2012.

[37] I. Simanova, M. Van Gerven, R. Oostenveld, and P. Hagoort, "Identifying object categories from event-related EEG: Toward decoding of conceptual representations," *PloS one*, vol. 5, no. 12, p. e14465, 2010.

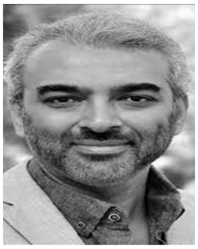
[38] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 701–711, 2005.

[39] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Comput. Biol. Med.*, vol. 34, no. 6, pp. 523–537, 2004.

[40] L. Wang and X. Fu, *Data Mining With Computational Intelligence*. New York, NY, USA: Springer, 2006.

RAHEEL ZAFAR received the master's degree from the University of Engineering & Technology, Taxila, Pakistan, in 2006. He is currently pursuing the Ph.D. degree with Universiti Teknologi Petronas, Malaysia. From 2007 to 2013, he served as a Lecturer with the Engineering Department, National University of Modern Languages, Islamabad, Pakistan.

SARAT C. DASS was an Assistant Professor and Associate Professor with Michigan State University, from 2000 to 2012. He is currently an Associate Professor with Universiti Teknologi Petronas, Malaysia.



AAMIR SAEED MALIK received the Ph.D. degree from the Gwangju Institute of Science & Technology, South Korea. He is currently an Associate Professor and the Ex-Director of the Biomedical Technology Group, one of the Mission Oriented Research with Universiti Teknologi Petronas, Malaysia.

NIDAL KAMEL received the Ph.D. degree (Hons.) from the Technical University of Gdansk, Poland, in 1993. He is currently an Associate Professor with Universiti Teknologi Petronas, Malaysia, and he is the Leader of Neuro-Signal Processing Group with the Center of Image and Signal Intelligent Research.

M. JAVVAD UR REHMAN received the master's degree from Quai-e-Azam University, Islamabad, Pakistan, in 2010. He is currently pursuing the Ph.D. degree with the Universiti Teknologi Petronas, Malaysia. From 2010 to 2014, he served as a Lecturer with the Engineering Department, National University of Modern Languages, Islamabad, Pakistan.

RANA FAYYAZ AHMAD received the master's degree from the National University of Science & Technology, Islamabad, Pakistan, in 2009. He is currently pursuing the Ph.D. degree with Universiti Teknologi Petronas, Malaysia.

JAFRI MALIN ABDULLAH is currently a Director or Founding Director of the Service Center and Neuroscience Research, Universiti Sains Malaysia (P3Neuro) the first in Southeast Asia and Founding Head and the former Head of the Department of Neuroscience, Health Campus, Universiti Sains Malaysia.

FARUQUE REZA received the Ph.D. degree from Hokkaido University, Japan, in 2005. He holds a post-doctoral position with Nagoya University, Japan, from 2006 to 2009. He is currently an Associate Professor with the Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia.

• • •