Neurocomputing 417 (2020) 558–584

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A novel approach for classification of mental tasks using multiview ensemble learning (MEL)

A. Gupta [a,e], R.U. Khan [b], V.K. Singh [a], M. Tanveer [b,*], D. Kumar [c], A. Chakraborti [a], R.B. Pachori [d]

[a] School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India
[b] Discipline of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India
[c] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India
[d] Discipline of Electrical Engineering, Indian Institute of Technology Indore, Simrol, Indore 453552, India
[e] Central Electronics Engineering Research Institute Pilani, Rajasthan 333031, India

ARTICLE INFO

Article history:
Received 9 April 2019
Revised 2 May 2020
Accepted 17 July 2020
Available online 1 August 2020
Communicated by Zhu Yu

Keywords:
Multiview learning
Brain computer interface (BCI)
Electroencephalography (EEG)
Mental tasks classification
Wavelet transform (WT)
Empirical mode decomposition (EMD)
Fuzzy C-means
Feature coding
Support vector machine (SVM)

ABSTRACT

Brain-computer interface (BCI) is a domain, in which a person can send information without using any exterior nerve or muscles, just using their brain signal, called electroencephalography (EEG) signal. Multiview learning or data integration or data fusion from a different set of features is an emerging way in machine learning to improve the generalized performance by considering the knowledge with multiple views. Multiview learning has made rapid progress and development in recent years and is also facing many new challenges. This method can be used in the BCI domain, as the meaningful representation of the EEG signal in plenty of ways. This study utilized the multiview ensemble learning (MEL) approach for the binary classification of five mental tasks on the six subjects individually. In this study, we used a well-known EEG database (Keirn and Aunon database). The EEG signal has been decomposed using by methods i.e wavelet transform (WT), empirical mode decomposition (EMD), empirical wavelet transform (EWT), and fuzzy C-means followed by EWT (FEWT). After that, the feature coding technique is applied using parametric feature formation from the decomposed signal. Hence, we had four views to learn four same type of independent base classifiers and predictions are made in an ensemble manner. The study is performed independently with three types of base classifiers, i.e., K-nearest neighbor (KNN), support vector machine (SVM) with linear and non-linear kernels The performance validation of the ten combinations of mental tasks was performed by three MEL based classifiers, i.e., K-nearest neighbor (KNN), support vector machine (SVM) with linear and non-linear kernels. For reliability of the obtained results of the classifiers, 10-fold cross-validation was used. The proposed algorithm shows a promising accuracy of 80% to 100% for binary pair-wise classification of mental tasks.
© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Electroencephalogram (EEG) signals are widely used by researchers to record brain activities through the EEG recording machine for the multiple applications of the brain-computer interface (BCI). The electrical activity of the brain can be captured from different EEG sensors and represented in various domains; it is employed for the classification of different EEG signals related to different mental tasks in EEG based BCI systems. It is a highly complex problem for the interpretation of brain activities using electrical signals. Five classifiers were used by Khare et al. [1] to classify four different mental tasks by using resilent back propagation algorithm. Five mental activities were classified by Garrett et al. [2] using support vector machine (SVM). A system developed by Kerkeni et al. [3] based on multi-layer perceptron (MLP) was used to recognized whether the person is awake or sleeping. The classification of EEG signal is done by Aljazaery et al. [4] using quantum neural network. A neural network used by Nandish et al. [5] to classify EEG signals. Recently, Richhariya and Tanveer [6] developed a universum based SVM model for the classification of healthy and seizure EEG signals.

Palaniappan [7] used some additional frequency band (24–37) Hertz (Hz) and calculated asymmetry and spectral power based features, this method gives some satisfactory performance of classification. Zhang et al. [8] used similar technique of feature

* Corresponding author.
E-mail addresses: akshanshgupta@jnu.ac.in (A. Gupta), riyaz@iiti.ac.in (R.U. Khan), vinod.acear@gmail.com (V.K. Singh), mtanveer@iiti.ac.in (M. Tanveer), dhirendrabhu08@gmail.com (D. Kumar), anirban@jnu.ac.in (A. Chakraborti), pachori@iiti.ac.in (R.B. Pachori).

https://doi.org/10.1016/j.neucom.2020.07.050
0925-2312/© 2020 Elsevier B.V. All rights reserved.

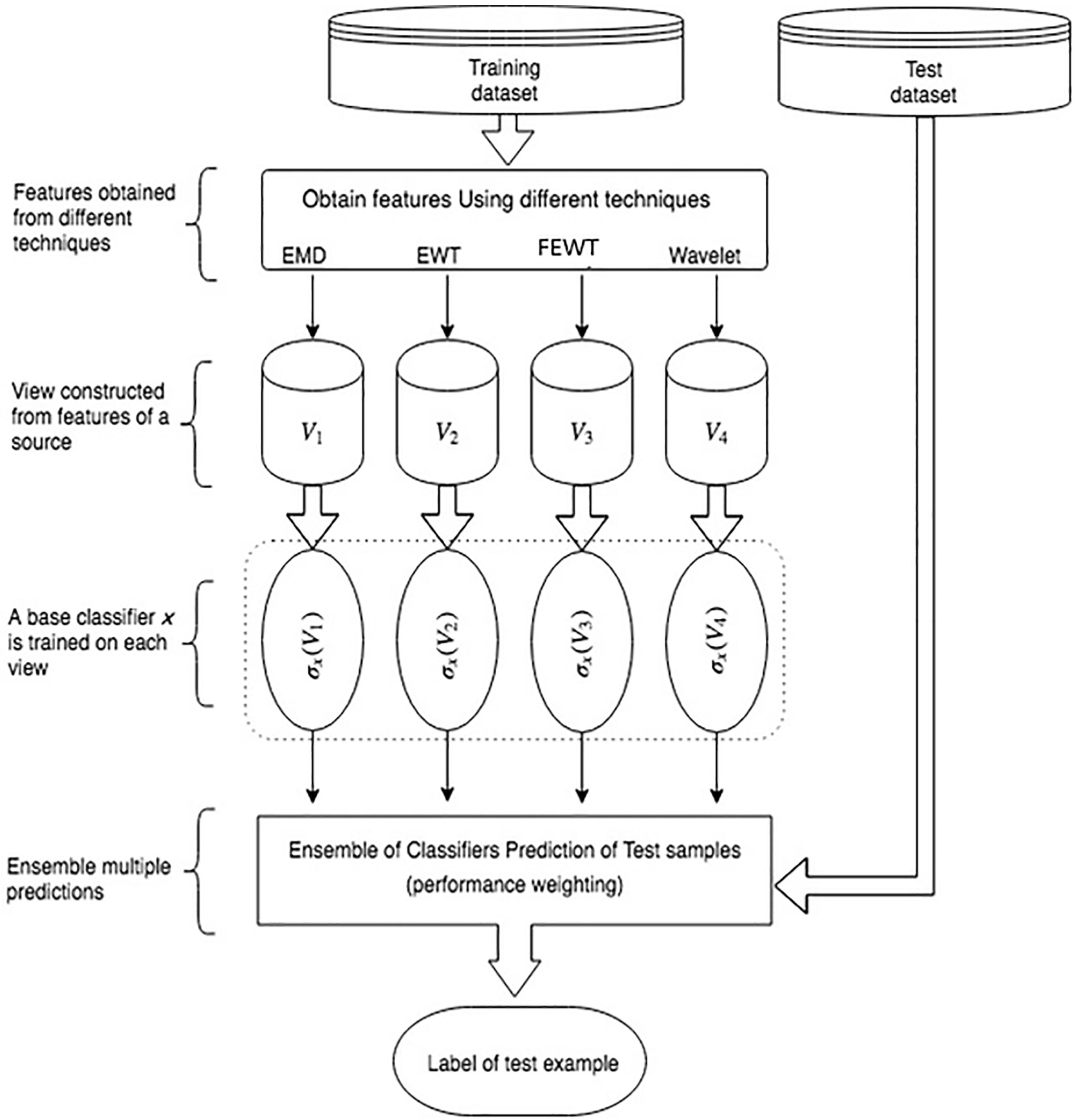

Fig. 1. Computational prediction strategy using MEL approach on 4 views.

extraction but different frequency band (40–100 Hz) to get satisfactory performance. Ameri et al. [9] classify several mental tasks by introducing a dictionary comprising of power spectral density and common spatial pattern (CSP) algorithm. Gupta et al. [10] have utilized power spectral density with feature selection to classified binary mental task. Gupta et al. [11] used empirical mode decomposition (EMD) and wavelet transform (WT) for classification of mental tasks by utilizing feature selection methods. Tanveer et al. [12,13] developed authomated models for the classification of EEG signals using flexible analytic wavelet transform (FAWT) based on Hjorth parameters and entropy based features. Hariharan et al. [14] used mean square root of standard deviation of signal as feature after applying Stockwell transform for classification performance measure. Siuly et al. [15] introduces a method in which cross correlation based feature extraction was utilized, in which cross correlation between different channels were calculated by keeping one signal as reference signal.

In current years, multiview learning methods have been evolved for data analysis by considering the different views. These views can be obtained from different feature subsets. Multiview learning mainly deals with the problems concerned with machine learning in which data is represented by the integration of different feature sets. In multiview learning, the use of consistency among different views aims to get better accuracy of performance. In

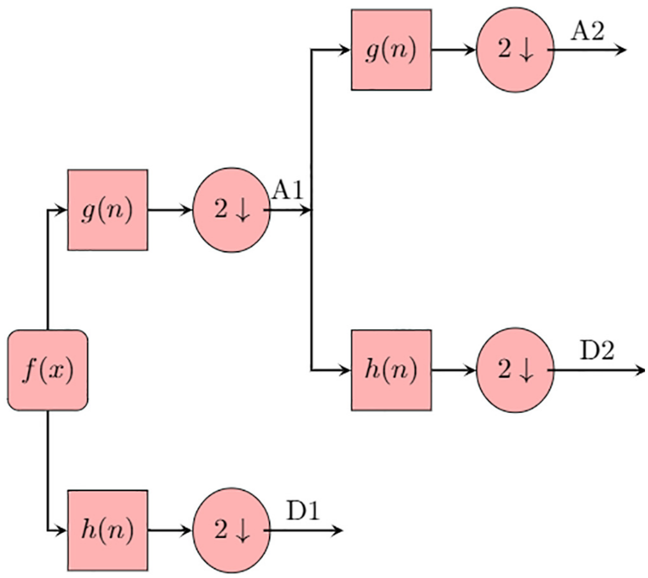

Fig. 2. Detailed and approximate coefficients of a signal.

multiview learning, multiple classifiers are trained on lower-dimensional feature subsets, which reduce the complexity of the trained model significantly, and variance error is lowered. After that, these multiple trained classifiers make predictions in an ensemble manner, and this approach reduces the bias error. Hence, the multiview ensemble learning (MEL) model reduces the overall error of the model by reducing both bias and variance parts. The classification of existing multiview learning methods can be done in the following three groups [16]: multiple kernel learning, co-training learning, and subspace learning. For multiview learning, co-training is one of the pioneer methods. In co-training learning, the maximization of mutual arrangement is done by training alternatively on two different views of data [17–19]. Some of them are co-expectation maximization (co-EM) [20], robust co-training [21] and co-testing [22], which belong to co-training learning. Multiple kernel learning combines the kernels which are either linear or nonlinear to improve the performance by exploiting the kernels which naturally correspond to distinct views [19,23]. Multiview Laplacian SVMs [24], sparse multiview support vector machines (SVMs) [25], multiview twin SVMs (TSVMs) [26] and multiview Laplacian TSVMs [27] are some illustrative algorithms for multiple kernel learning. In subspace learning algorithm a latent subspace is shared by multiple views with the assumption that the input views are created from the latent subspace [28]. Realization of this algorithm is done on the basis of maximize entropy discrimination (MED) [29].

In addition to the latest proposed multiview learning approaches, certain detailed algorithm of multiview learning are sequentially put forward for definite machine learning tasks. They can be summarized as multi-task multiview learning [30,31], multiview semi-supervised learning [24,27], multiview dimensionality reduction [32,33], multiview transfer learning [34,35], multiview discriminant analysis [36,37] and multiview clustering [38,39].

The main objective of this paper is to apply multiview learning for mental task classification using four different sets of features. The features extracted from different algorithms are allocated to separate and independent views. A base classifier is trained independently on each view and prediction is made in an ensemble manner on test dataset. The performance of various classification algorithms is dependent on the distribution, size and dimension of the dataset. So, we have carried out our study separately using three independent base classification algorithms i.e., k-nearest neighbor (KNN), linear SVM and nonlinear SVM using radial basis function (RBF) kernel. The performance of our proposed MEL based classification approach is reported in terms of accuracy (acc), the area under the curve (AUC), sensitivity (sn), specificity (sp) and Mathews correlation coefficient (MCC) and compared to the existing methods.

The SVM and KNN classifiers have been widely used for automated classification of various biomedical signals. For example, the SVM has been used for automated identification of alcoholism from EEG signals [40], automated identification of stress using EEG signals [41], and automated detection of epilepsy based on EEG signals [42]. On the other hand, KNN has been studied for automated classification of sleep stages from EEG signals [43], automated identification of epileptic seizures from EEG signals [44], and automated detection of various cognitive states from EEG signals [45]. These studies have provided a motivation for studying SVM and KNN classifiers for automated classification of the EEG signals in our paper.

To the best of our knowledge, multiview learning has not done so far in the field of mental task classification.

The paper is organized as follows. Section 2 explains MEL. Section 3 explains the methodology used. The proposed method is elaborated in the SubSection 3.1. In Section 4 detail of creating multiview is given, whereas the brief description of classifiers, which have used in this work, is given in the Section 5. Section 6 explains the classification results obtained. Finally, Section 7 describes conclusion and future work.

2. Multiview ensemble learning (MEL)

In MEL, for each view of dataset different classification algorithms are applied to build classifiers of the corresponding view. The trained classifier individually predicts the class label of the test sample of corresponding views. Finally, for obtaining the class labels of test sample all the predictions are the ensemble. It is known that non-relevant and redundant features are utilized in MEL. MEL performance will degrade if the learning algorithms are not matched to the view. Therefore, to ensure the effective performance of MEL complementary principles and consensus are considered while selecting the views, as follows:

1. *Complementary principle:* The information contained in one view of data may not be present in other views, called complementary information. Hence, for improving the performance of

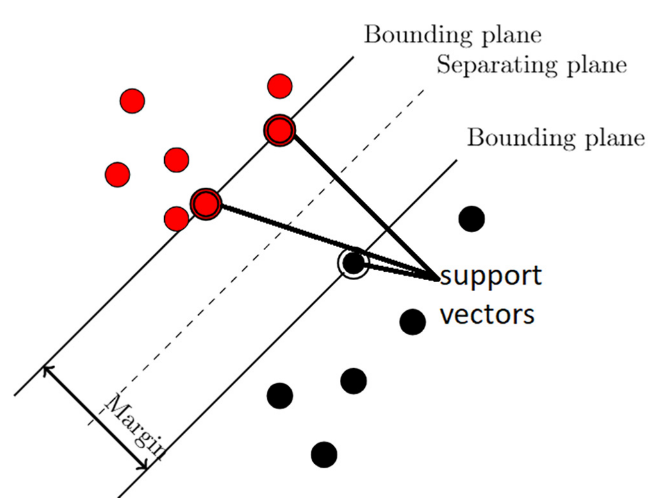

Fig. 3. Geometrical interpretation of SVM.

Table 1
AUROC of classification of 2-class using different classifiers.

Subject	classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	0.8974	0.8486	0.9094	0.9431	0.7746	0.9172	0.8007	0.9552	0.8753	0.9834
	MEL-SVM(L)	0.8546	0.7635	0.9124	0.9225	0.6791	0.929	0.8118	0.9421	0.7888	0.9794
	MEL-SVM(R)	0.9453	0.8425	0.941	0.96	0.6812	0.9468	0.836	0.9688	0.8712	0.9842
Sub2	MEL-KNN	0.8548	0.8476	0.883	0.9906	0.7312	0.7572	0.9578	0.839	0.9766	0.9291
	MEL-SVM(L)	0.8878	0.7951	0.9158	0.9958	0.66	0.8023	0.9749	0.811	0.9921	0.9546
	MEL-SVM(R)	0.8975	0.8691	0.9331	0.9944	0.7376	0.8988	0.9816	0.9327	0.9895	0.9511
Sub3	MEL-KNN	0.4857	0.6482	0.7894	0.6049	0.7259	0.7858	0.6223	0.8596	0.693	0.8597
	MEL-SVM(L)	0.5711	0.6964	0.8675	0.6163	0.7148	0.7782	0.5656	0.8662	0.6651	0.7864
	MEL-SVM(R)	0.6699	0.7016	0.8786	0.6454	0.7326	0.8127	0.6458	0.8682	0.7139	0.8364
Sub5	MEL-KNN	0.8212	0.8658	0.8586	0.8687	0.7045	0.7383	0.8323	0.8197	0.8602	0.8157
	MEL-SVM(L)	0.7988	0.8611	0.8418	0.8672	0.7453	0.6853	0.7604	0.7675	0.8917	0.7636
	MEL-SVM(R)	0.8267	0.8746	0.8617	0.8702	0.7383	0.7468	0.7792	0.7764	0.9233	0.8069
Sub6	MEL-KNN	0.7924	0.6592	0.8484	0.8491	0.7145	0.8873	0.82	0.7763	0.8055	0.7796
	MEL-SVM(L)	0.8247	0.6271	0.8677	0.8164	0.7223	0.9517	0.8222	0.8111	0.7469	0.8093
	MEL-SVM(R)	0.8624	0.659	0.8825	0.8553	0.7587	0.9633	0.8541	0.9078	0.788	0.844
Sub7	MEL-KNN	0.8979	0.8329	0.9722	1	0.8225	0.8335	0.9913	0.811	0.9916	0.8693
	MEL-SVM(L)	0.9301	0.8763	0.9435	0.9958	0.7726	0.8095	0.9908	0.9171	0.9958	0.8767
	MEL-SVM(R)	0.886	0.8633	0.9574	1	0.8427	0.9738	0.9955	0.9781	0.9944	0.8703

Note: B-Baseline Task, C–Counting Task, L-Letter Composition Task, R-Rotation Task, M-multiplication Task.

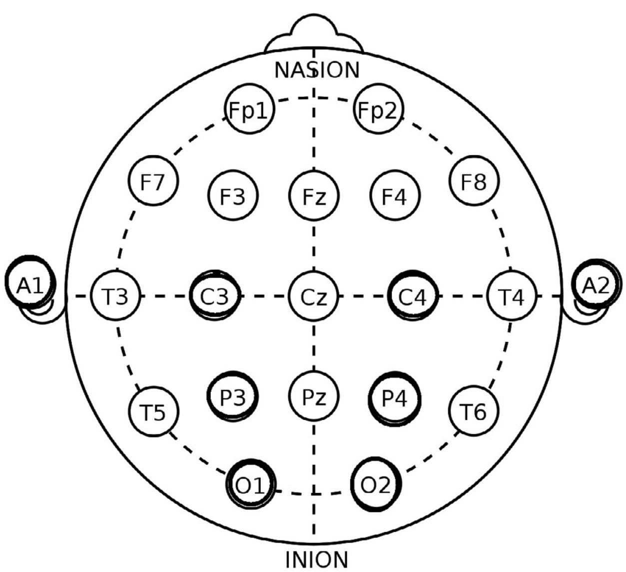

Fig. 4. 10–20 standard for EEG recording.

the learning algorithms the information from different views can be employed.
2. *Consensus principle:* Objective of this principle is to limit the contradiction among views of the unlabeled data.

The following three steps of MEL are as follows: (1) View construction, (2) View evaluation, and (3) View ensemble, which are defined as follows:

1. **View construction:** Various optimization-based strategies have been proposed to efficiently partition the feature set into multiple feature-subsets (views) of lower dimension [46]. However, this study does not require partitioning of feature set. The features extracted by a method are considered as a view, and in total, four methods have been used to extract features from the EEG data. Hence, there are four views available to build a ensemble learning model.
2. **View evaluation:** The performance of MEL classifiers depends on the views. Therefore, each view is evaluated by machine learning algorithms in terms of the accuracy on the same view [16].
3. **View ensemble:** The label of the test sample is the ensemble of the labels predicted by trained classifiers on the given views [47]. Our study utilized the performance weighting based ensemble approach, where the accuracy of the classifier trained on a view of the validation set is used as the weight of classifier to carry out ensemble prediction [47].

3. Methodology

If we combine all features obtained from the feature extraction techniques, which have been explained in subSection 4.1, then the dimensionality of data increased, whereas, limited number of samples were available. This may cause over-fitting to the classification model. To overcome this challenge and enhance generalization of our model we trained same type of base classifiers on each subset of features in an independent and parallel manner, where, a subset of features is known as a view. These multiple trained homogeneous base learners made the multiple independent predictions about an unlabeled test example and the final prediction is the ensemble of these predictions using the performance weighting method [16,48]. This approach of ensemble prediction by a classifier trained multiple views of the data set is known as MEL.

For the study, the feature set of a particular feature extraction method was considered to construct a view. Thus, we have four views each belonging to a feature extraction methods i.e., EMD, EWT, fuzzy C-means followed by EWT (FEWT), and WT, to train four homogeneous base learners and make the ensemble of predictions on the test dataset. To find a suitable base learner for our dataset, we carried out our independent study with three base learning algorithms, i.e., SVM with linear and RBF kernel and KNN and robust ensemble classification model. The grid search approach [49] was used to select the optimal set of parameters for the classification models i.e., SVM (linear/RBF) and KNN and 10-fold cross-validation was used to minimize prediction error. The results are summarised in results and discussion section part of this paper.

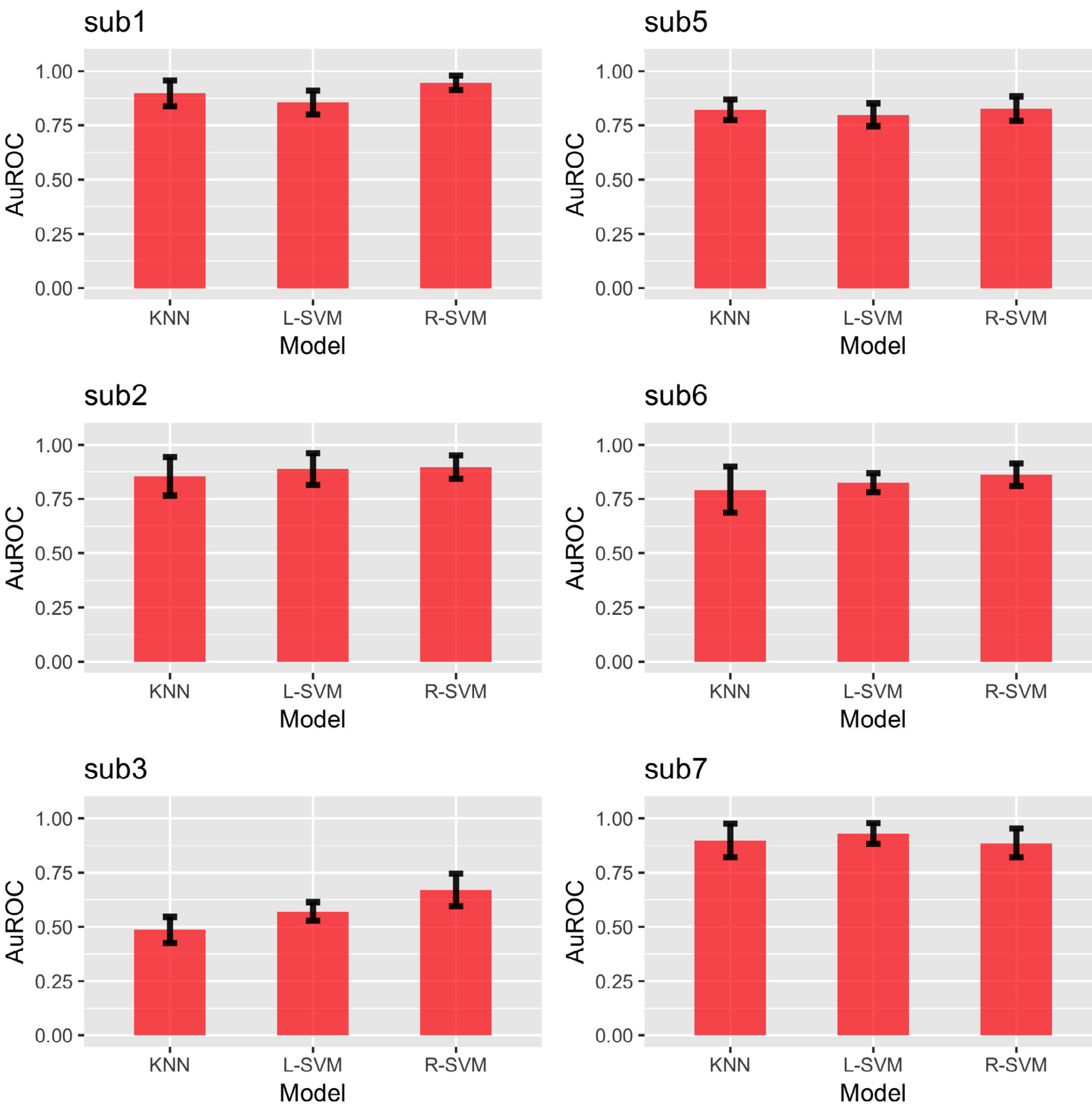

Fig. 5. AUROC for baseline vs count combination of all subjects using different classifiers.

3.1. Proposed approach

3.1.1. MEL algorithm:

The datasets used for MEL from k independent views can be represented as $A_k^s = \left\{ D_v^s = \left[d_{v,1}^s, d_{v,2}^s, \ldots, d_{v,n_s}^s \right] \in R^{dim_v \times n_s} \right\}_{v=1}^{k}$, where dim_v is the number of features in the v^{th} view and n_s is the number of samples. We used 90% data for training and remaining 10% for testing. The label of the test sample was the performance weighted ensemble of the labels predicted by individual classifiers trained on each view [48,50]. The detailed methodology is graphically represented in the Fig. 1.

3.1.2. Flowchart of MEL strategy

The following steps were adopted in MEL based prediction of mental tasks from the EEG data. The features obtained by a method were considered as a view. In our study, we have used four methods to extract features from the EEG signals. Therefore, we had four views, and each view was used to train a classifier. At last, the performance-weighting based ensemble prediction of the test example was made by four independent classifiers in Fig. 1 clearly represents this methodology.

4. Methods used for creating multiview

4.1. Feature extraction techniques

Features from EEG signal can be extracted in many ways. We used WT and EMD for decomposition of EEG signal and two other techniques Fuzzy C-mean and feature coding to estimate statistical parameters i.e. features from EEG signal. The brief introduction of statistical parameters is discussed below:

4.1.1. Wavelet transform (WT)

WT is a multi-resolution tool, which is used to analyze the signal in both frequency and time domain simultaneously by using a

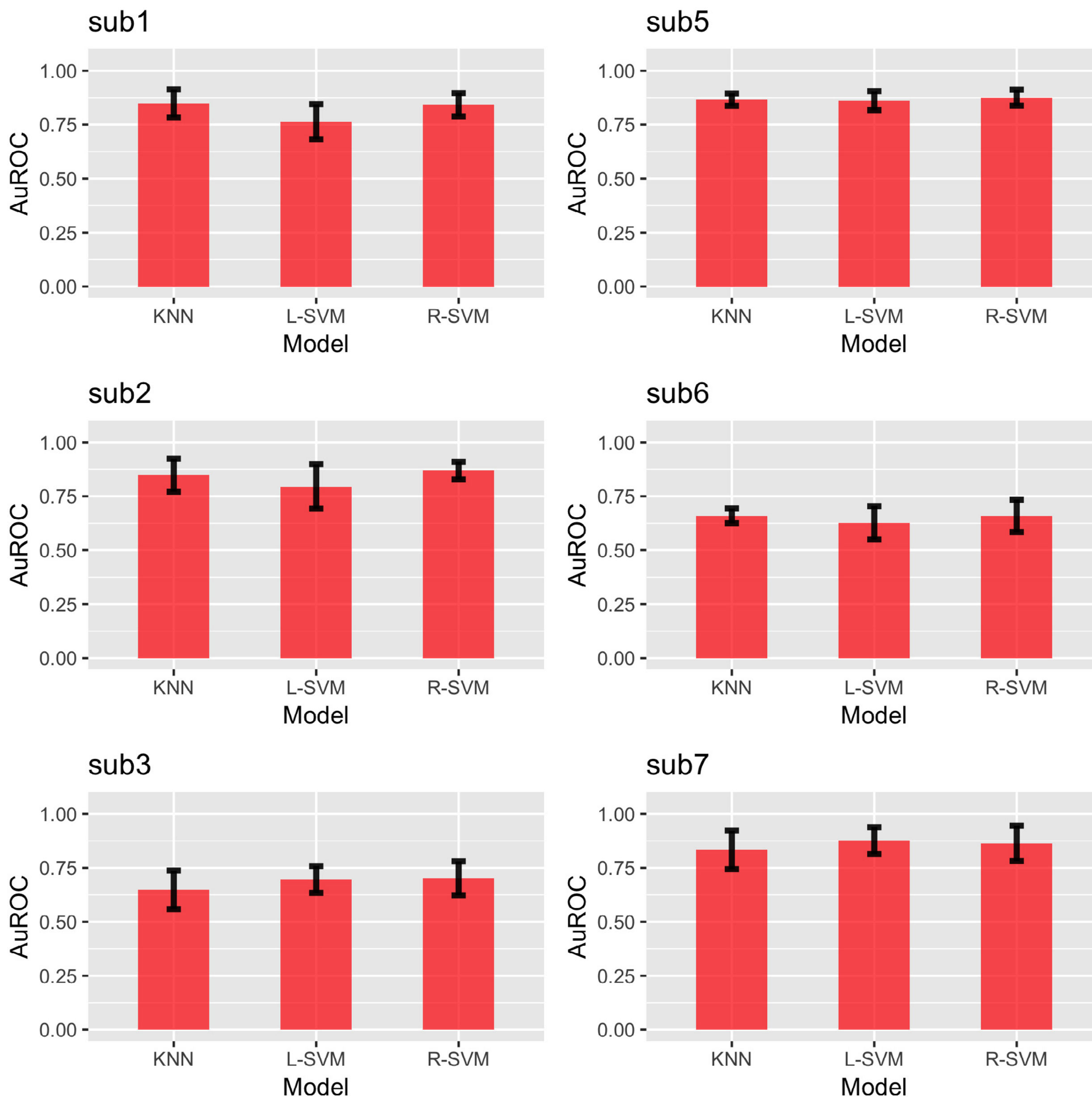

Fig. 6. AUROC for baseline vs letter combination of all subjects using different classifiers.

window of variable length. 1-D signal $f(x)$ can be decomposed by discrete wavelet transform (DWT) in terms of scaling function $\phi(x)$ and a shifted and dilated mother wavelet $\psi(x)$ described by Mallat [51]. The decomposition of signal is clearly shown in Fig. 2. in terms of approximation coefficients A and the detail coefficient D. The detailed information is produced by high pass filter; while approximate information is produced by low pass filter. The signal $f(x)$ is represented by:

$$f(x) = \sum_{l} s_{j_0,l}\phi_{j_0,l}(x) + \sum_{j=j_0}^{\infty}\sum_{l} d_{j,l}\psi_{j,l}(x), \tag{1}$$

where j_0 is an arbitrary starting scale, $l = 0, 1, 2, \ldots, 2^{j-1}$, $s_{j_0,l}$ and $d_{j,l}$ are scaling and wavelet coefficients respectively. The scaling coefficients and wavelet coefficients can be calculated as follows:

$$s_{j,l}(x) = \langle f(x), \phi_{j,l}(x)\rangle, l = 0, 1, 2, \ldots, 2^{j-1}, \tag{2}$$

$$d_{j,l}(x) = \langle f(x), \psi_{j,l}(x)\rangle, l = 0, 1, 2, \ldots, 2^{j-1}. \tag{3}$$

The scaling function and wavelet basis function at j are respectively given below:

$$\phi(x)_{j,k} = \sum_{n} h(n)\phi(x)_{j+1,n}, \tag{4}$$

$$\psi(x)_{j,k} = \sum_{n} g(n)\psi(x)_{j+1,n}, \tag{5}$$

where $g(n)$ is high pass filter, $h(n)$ is a low pass filter.

4.1.2. Empirical mode decomposition (EMD)

If it is assumed that any signal is composed of sequence of different intrinsic oscillation modes, then it can be decomposed into different intrinsic mode functions (IMFs) using EMD algorithm. An IMF is a function, which fulfills the underlying conditions.

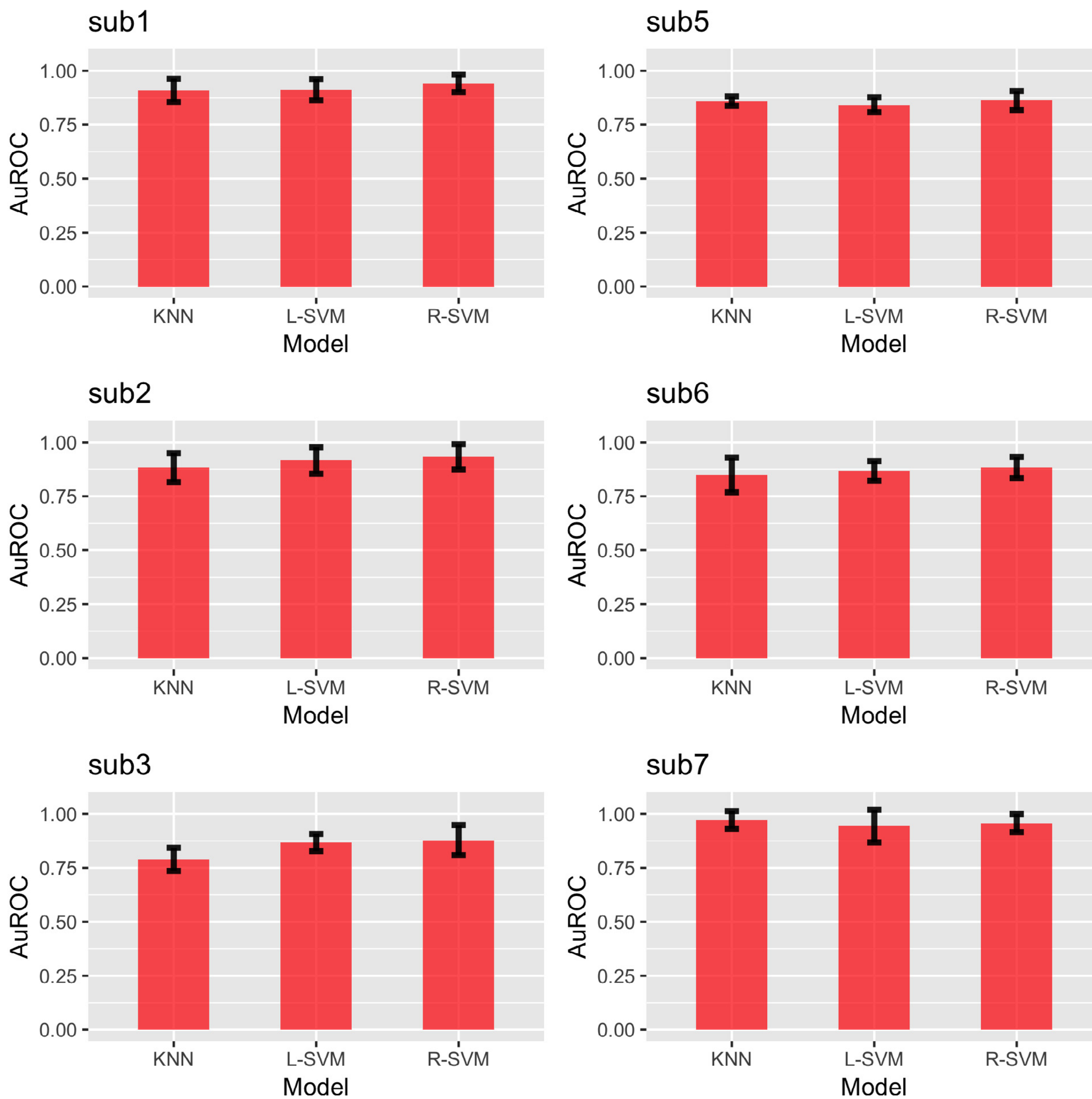

Fig. 7. AUROC for baseline vs math combination of all subjects using different classifiers.

1. There should be either equal, or differ at most by one, number of maxima and minima.
2. The envelope defined by local minima and maxima should have zero mean value.

The algorithm of EMD for the given incoming signal $f(x)$ is based on sifting process can be summarized as follows [52]:

1. Identify all local minima and maxima for given signal $f(x)$.
2. For calculating the upper envelope, all the local maxima points of signal are connected using a cubic spline.
3. For lower envelope, repeat the same for local minima points of signal.
4. Calculate the mean of the two envelopes, let it be m_1.
5. The signal is update to, $f(x) = f(x) - m_1$.
6. Steps 1 to 5 should be repeated, until $f(x)$ is considered to be an IMF according to the definition expressed previously by considering $f(x)$ as the input signal.
7. Then obtain the residue r_1 by subtraction of the first IMF (IMF_1) from $f(x)$ i.e., $r_1 = f(x) - IMF_1$. For the next iteration, the residue of this step becomes the signal $f(x)$.
8. To get all the IMFs of the signal steps 1 to 7 are iterated on the residual $r_j = 1, 2, 3, \ldots, n$.

The technique ends when the residual r_j is either a constant value or a function with a solitary maxima (minima). As a result of EMD process n IMFs and a single residue signal r_n is produced. The n extracted IMFs and the residue terms can be used to reconstruct the original signal $f(x)$ as follows:

$$f(x) = \sum_{j=1}^{n} IMF_j + r_n. \tag{6}$$

To get the IMFs of the signal, openly accessible EMD tool kit for Matlab is used. Faster oscillation modes of the signal are captured by lower order IMFs while slower oscillation modes are captured by higher order IMFs. It has been found during the decomposing

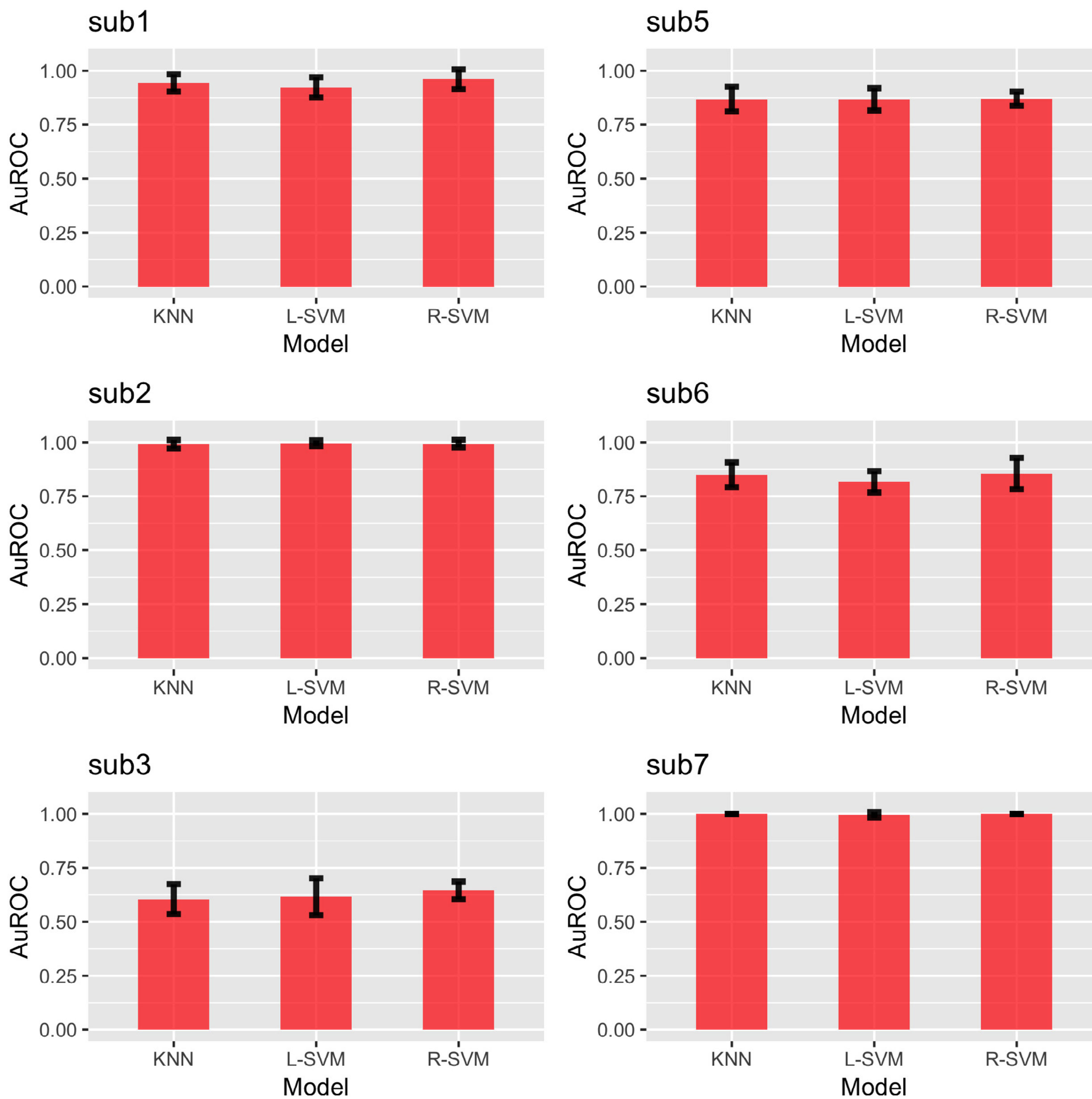

Fig. 8. AUROC for baseline vs rot combination of all subjects using different classifiers.

of the signal using EMD, most of the segments decompose into 4 or more IMFs [53]. Therefore, to preserve consistency, we have utilized four levels of IMF decomposition to build the feature vector of a similar size from a given sample.

4.1.3. Empirical wavelet transform

EEG signal has non-stationary and non-linear nature [54], in few past years, adaptive filter based EMD methods [55] and fixed basis function based WT [56,57] were used for studying the non-linear and non-stationary nature of EEG signals. The lack of mathematical theory is a major concern of EMD method [58]. The properties of these two methods can be combined by recently proposed method by Gilles [58] in which a new adaptive basis transform known as EWT is used to extract the mode of amplitude modulated-frequency modulated (AM-FM) signal. In these methods, a group of adaptive (empirical) wavelets of the signal is build. which is processed in the same manner as a set of band pass filters is formed in Fourier spectrum. The plan to accomplish the adaptability is the dependency of filter's supports on the area of the information in the spectrum of the signal [58].

Let ω represent the frequency which belongs to a segmented of N continuous segment, Fourier support $[0, \pi]$. The limit between the segment ($\omega_0 = 0$ and $\omega_N = \pi$) and $\lambda_n = [\omega_{n-1}/\omega_n]$ represents a segment such that $\bigcap_{n=1}^{N} \lambda_n = [0, \pi]$. Gilles [58] in his research work has assumed that each segment is having transition phase which is centered around ω_n having width of $2\tau_n$.

The ideas given by Little-Wood-Paley and Meyer's wavelets [59] can be utilized to define the empirical wavelet as a band pass filter of each λ_n. The empirical scaling function can be described as follows:

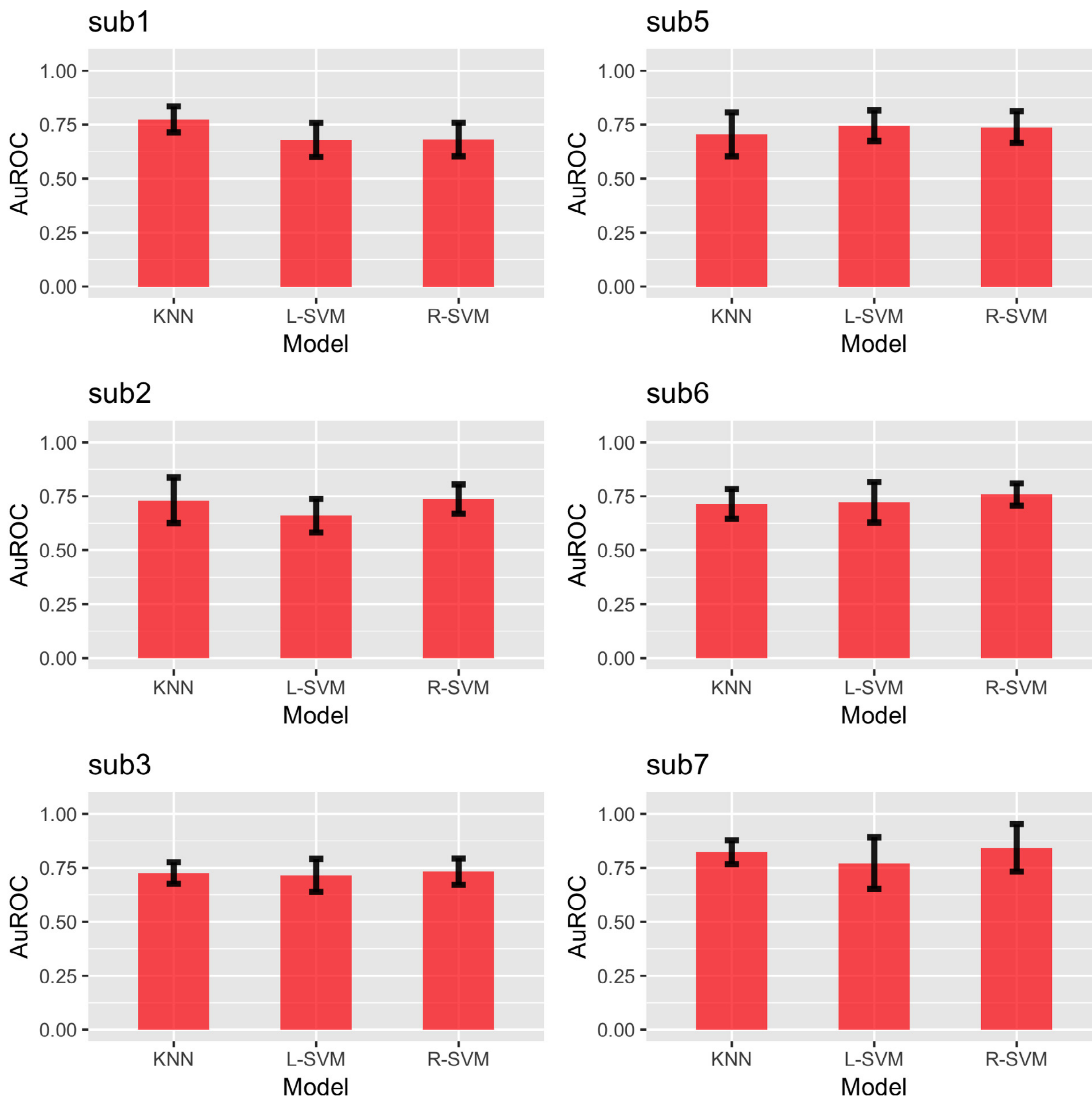

Fig. 9. AUROC for count vs letter combination of all subjects using different classifiers.

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leqslant \omega_n - \tau_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n - \tau_n)\right)\right] & \text{if } \omega_n - \tau_n \leqslant |\omega| \leqslant \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The empirical wavelets are given by equation as below:

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \leqslant |\omega| \leqslant \omega_{n+1} + \tau_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}(|\omega| - \omega_{n+1} - \tau_{n+1})\right)\right] & \text{if } \omega_{n+1} - \tau_{n+1} \leqslant |\omega| \leqslant \omega_{n+1} + \tau_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n - \tau_n)\right)\right] & \text{if } \omega_n - \tau_n \leqslant |\omega| \leqslant \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

EWT of the signal $f(t)$. $W_f^{\epsilon}(n,t)$ is described in the same manner as that of classical WT [58]. The detailed coefficient is given by:

$$W_f^{\epsilon}(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau - t)}d\tau, \tag{9}$$

$$W_f^{\epsilon}(n,t) = \langle f, \psi_n \rangle = \left(\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)}\right)^V, \tag{10}$$

where $\langle\rangle$ represents the inner product. In similar way, the approximate coefficient is given by,

$$W_f^{\epsilon}(0,t) = \langle f, \phi_1 \rangle = \int f(\tau)\overline{\phi_1(\tau - t)}d\tau, \tag{11}$$

$$W_f^{\epsilon}(0,t) = \langle f, \phi_1 \rangle = \left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right)^V. \tag{12}$$

The original signal $f(t)$ can be reconstructed by the following equations:

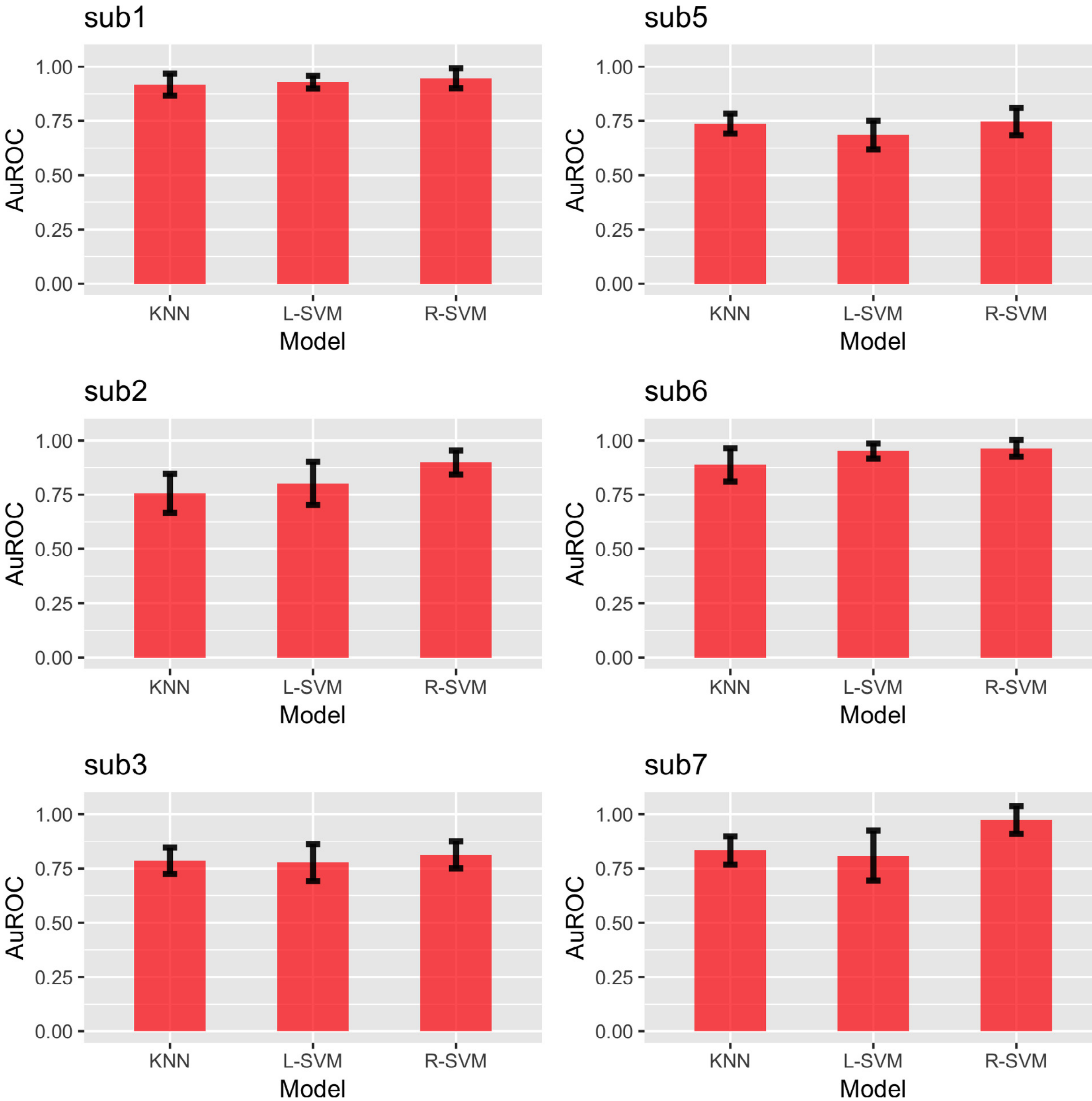

Fig. 10. AUROC for count vs math combination of all subjects using different classifiers.

$$f(t) = W_f^{\epsilon}(0,t)\bigstar\phi_1(t) + \sum_{n=1}^{N} W_f^{\epsilon}(n,t)\bigstar\psi_n(t), \tag{13}$$

$$f(t) = \left(\hat{W}_f^{\epsilon}(0,\omega)\bigstar\hat{\phi}_1(\omega) + \sum_{n=1}^{N} \hat{W}_f^{\epsilon}(n,\omega)\bigstar\hat{\psi}_n(\omega)\right)^{V}. \tag{14}$$

where $\bigstar$ represents convolution operation.

4.1.4. Fuzzy C-means

Fuzzy C-means algorithm [60] is a clustering technique based on fuzzy set theory which was developed by Zadeh [61] and is seen in various prospects by a few researchers such as Nguyen [62] and Tiwari et al. [63]. The main idea of Fuzzy C-means Clustering (FCM) is that, a single object can be grouped in more than one cluster on behalf of fuzzy membership value ($[0,1]$) instead of on the ground of crisp value $(0,1)$ as in k-means algorithm. Optimization of non linear problem for FCM can be described as follows:

$$\begin{cases} \text{Min } A_m(\mathbf{U},\mathbf{V};\mathbf{X}) = \sum_{j=1}^{p}\sum_{i=1}^{c}(u_{ij})^m d^2(x_j, v_i) \\ \text{such that } \sum_{i=1}^{c} u_{ij} = 1, 1 \leqslant j \leqslant p \\ 0 \leqslant u_{ij} \leqslant 1, 1 \leqslant j \leqslant p, 1 \leqslant i \leqslant c \\ 0 \leqslant \sum_{i=1}^{c}(u_{ij}) \leqslant p, \forall_i. \end{cases} \tag{15}$$

where $\mathbf{X}$ denotes the number of objects from 1 to p, c represents the number of clusters and m represents the fuzzifier constant. u_{ij} represents the degree estimation of membership of jth object which belongs to ith cluster. The centroid matrix and fuzzy partition is represented by $\mathbf{V}$ and $\mathbf{U} = (u_{ij})_{c\times p}$ respectively. $d^2(x_j, v_i)$ represents the Euclidean distance between j^{th} object and i^{th} centroid.

The fuzzy membership value can be updated for given subject after k iteration as follows:

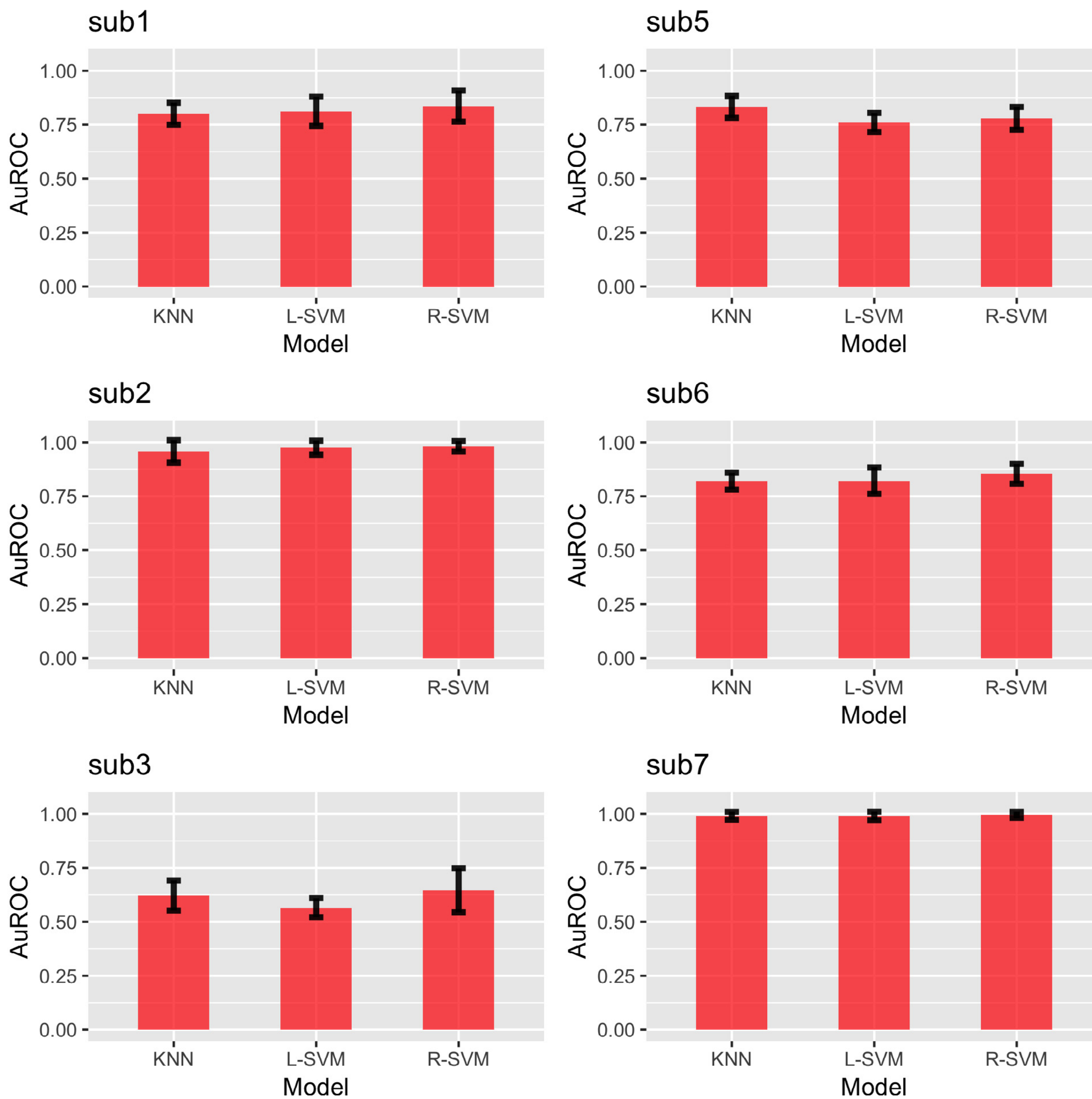

Fig. 11. AUROC for count vs rot combination of all subjects using different classifiers.

$$u_{ij}(k) = \frac{1}{\sum_{r=1}^{c}\left(\frac{d(x_j, v_i(k))}{d(x_j, v_r(k))}\right)^{\frac{2}{m-1}}}. \quad (16)$$

In similar way the updating of centroid can be done as follows [60]:

$$v_i(k+1) = \frac{\sum_{j=1}^{p} u_{ij}^m(k) x_j}{\sum_{j=1}^{p} u_{ij}^m(k)} \text{where} \leqslant i \leqslant c. \quad (17)$$

4.1.5. Feature coding

In the first phase of feature extraction modoule EEG signal is decomposed with help of either one the ways from WT, EWT, FEWT, and EMD. In final step, for representing every obtained segment more compactly, eight statistical or uncertainty parameters (root mean square, Lempel–Ziv complexity measure [64], Shannon entropy, central frequency, maximum frequency, variance, skewness, and kurtosis) were calculated.

After first phase of feature extraction modoule and eight statistical or uncertainty parameters calculation our four views were constructed, which can be summarized as follows:

1. WT and 8 features.
2. EWT and 8 features.
3. FEWT and 8 features.
4. EMD and 8 features.

5. Classification

We used three machine learning classifiers which are thoroughly described here. A classifier uses a set of independent variables i.e. features as an input according to which it will predict the equivalent class to which the given independent class belongs.

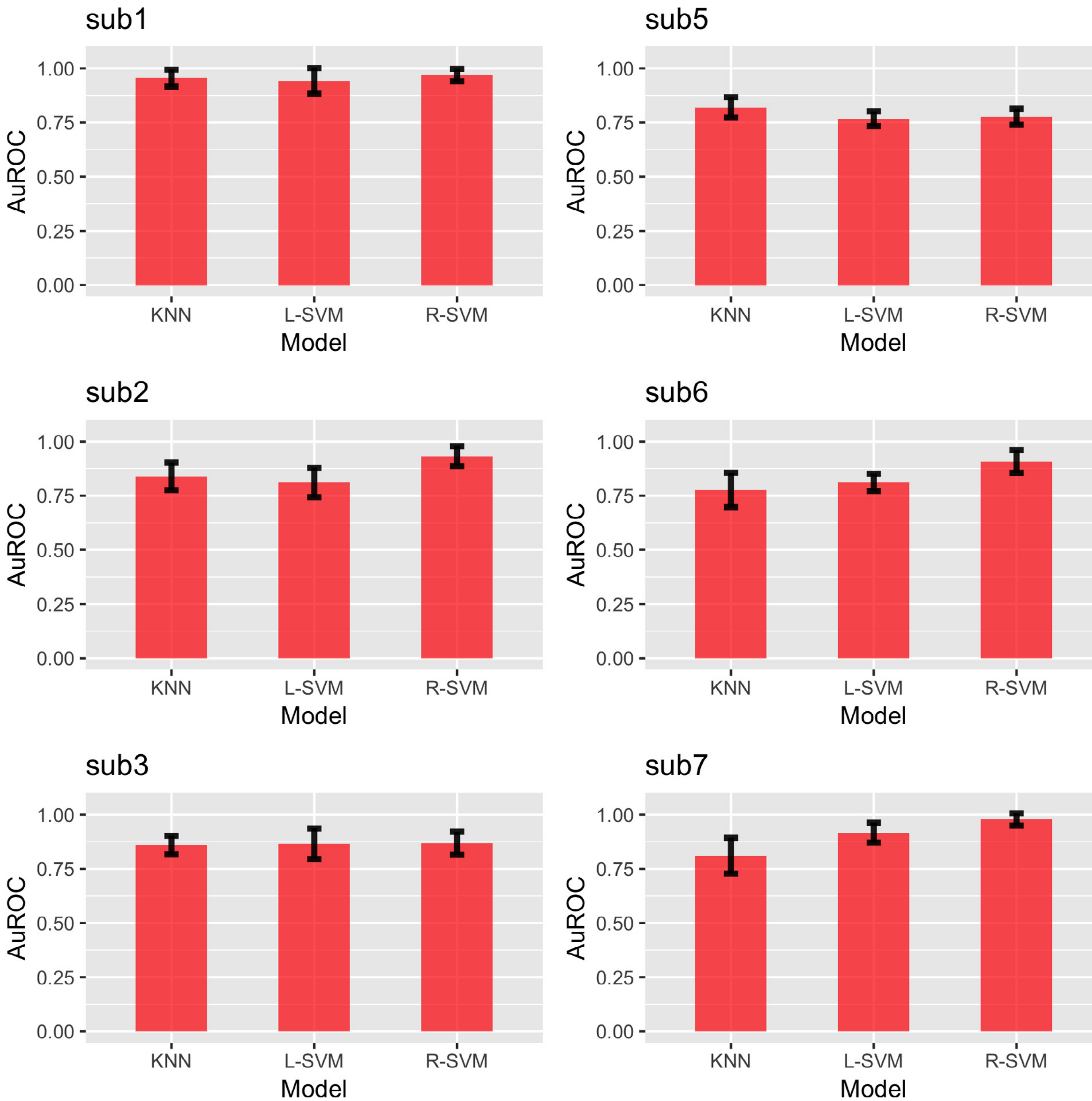

Fig. 12. AUROC for letter vs math combination of all subjects using different classifiers.

A classifier uses number of parameters which includes the data for training from training dataset. After training the classifier will be model as the association of classes and corresponding features and the unknown testing dataset will be identified accordingly. To validate the efficiency of the proposed study we have used the following classification algorithms:

5.1. Support vector machine (SVM)

SVM is classification method which is based on statistical theory. In pattern recognition tasks SVM shows excellent performance. Linear separating hyperplane is used by SVM for creating classifier with maximal margin. Suppose a problem of binary classification in which a linearly separable dataset X of l points in real dimensional space of features is denoted by the matrix $X \in R^{l\times n}$. The diagonal matrix $D \in R^{l\times l}$ having entries D_{ii} $+1$ or -1 represents the corresponding target class of each data point $X_i; i = 1, 2, \cdot, l$. For the above stated problem, SVM's linear soft margin algorithm is used to solve the given primal quadratic programming problem as follows: (QPP)[65].

$$\begin{aligned} &\min_{w,b}\ \tfrac{1}{2}w^T w + ae^T y \\ &\text{s.t.}\ \ D(Xw + eb) + y \geqslant e, y \geqslant 0e. \end{aligned} \tag{18}$$

Where a represents penalty parameter and y are some non negative slack variables. The hyperplane which is optimal can be expressed as follows:

$$d(x) = w^T x + b, \tag{19}$$

where x represents a test data point in R^n i.e. $x \in R^n$. Eq. (18) has large number of constraints, so usually it's dual is solved. The Wolfe solved this dual and is given as follows:

$$\begin{aligned} &\max_{\alpha}\ e^T\alpha - \tfrac{1}{2}\alpha^T DXX^T\alpha \\ &\text{s.t.} e^T D\alpha = 0 \quad 0e \leqslant \alpha \leqslant ae. \end{aligned} \tag{20}$$

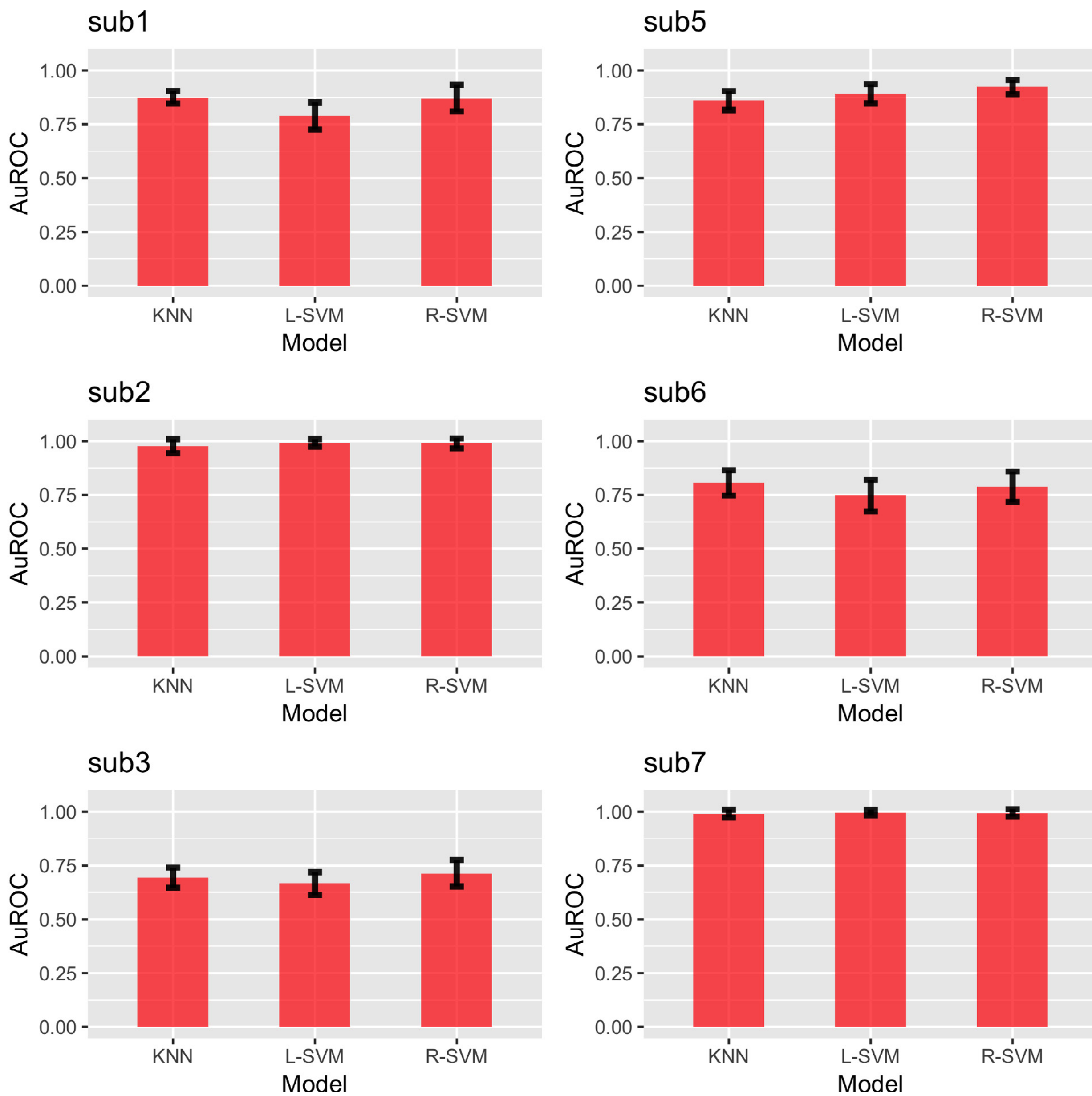

Fig. 13. AUROC for letter vs rot combination of all subjects using different classifiers.

α represents Lagrangian multipliers such that $\alpha \in R^l$. The parameters of this optimal separating hyperplane is same as that of (19) and are given by:

$$w = X^T D\alpha b = \frac{1}{N_{sv}} \sum_{i=1}^{N_{sv}} D_{ii} - X_i w, \tag{21}$$

where N_{sv} denotes the number of support vectors such that $0 \leqslant \alpha_i \leqslant a$. The hyperplane defined by (19) will lie midway between the bounding planes which are given as follows:

$$\mathbf{x}^T \mathbf{w}^{(1)} + b^{(1)} = 1, \quad \text{and} \quad \mathbf{x}^T \mathbf{w}^{(2)} + b^{(2)} = -1. \tag{22}$$

This will separate the two classes from one another by a margin of $\frac{2}{||w||}$. When a new data point is given it will be classified as $+1$ or -1 as the decision function $(w^T x + b)^*$ gives 1 or 0 respectively. The geometrical interpretation of this formulation is represented in Fig. 3. SVM has an important characteristic of extending to create non-linear boundaries in a relatively straightforward manner.*5.2.*

SVM with radial basis function (RBF) kernel SVM has been widely used for the classification of feature vectors of binary and multi-class problems [66,67]. In synchronization of BCI's, it has been successfully used, such types of classifiers are known as linear classifiers as it uses one or more hyperplanes. On the other hand, the non-linear decision boundaries based SVM are also possible by using kernel function:

$$K(x_i, x_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j). \tag{23}$$

By applying the kernel function, the non separable feature vectors are transformed into high dimensional feature vectors, where these features can be easily separable. Classification accuracy is increased in non-linear SVM due to more flexibly decision boundary in the vector space. The Gaussian or RBF is generally used as a kernel in BCI field applications. This function is given by;

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}. \tag{24}$$

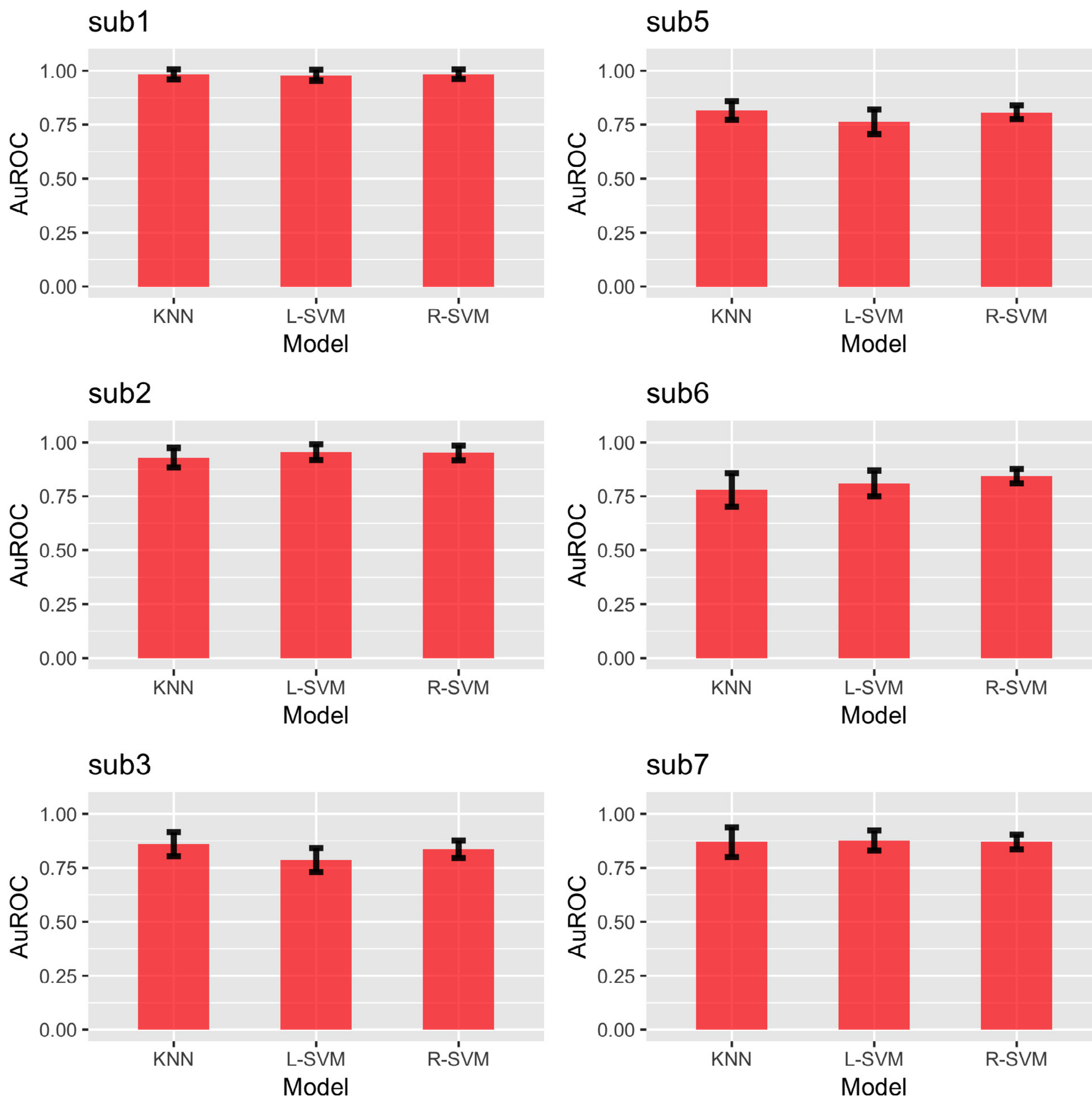

Fig. 14. AUROC for math vs rot combination of all subjects using different classifiers.

Considering the above kernel function, the dual formation will be given as follows:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \Phi(x_i)^T\Phi(x_j) \\ \text{s.t.} \quad & 0 \leqslant \alpha_i \leqslant a, \forall i = 1, 2, \ldots, n. \\ & \sum_{i=1}^{n}\alpha_i y_i = 0, \end{aligned} \tag{25}$$

where n represents the total number of data points in the sample. The following equation will give the solution of two class linearly non-separable problem:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n}\alpha_i y_i \Phi(x_i)^T\Phi(x_j) + b\right). \tag{26}$$

5.3. K-nearest neighbor (KNN)

In machine learning, KNN is one of the most essential classification algorithms. Its algorithm is based on supervised learning domain and has application in data mining, instruction detection, and pattern recognition. It is based on the principle that the features corresponding to the different classes will usually form separate clusters in the feature space, while the close neighbors belong to the same class.

It is a non-parametric occurrence based classifier [68]. It simply collects all the samples of features in training data and does not learn from training data that is why it is also known as lazy learning method [69]. The samples collected in training data are used during the training phase. The nearest neighbor estimation algorithm is used in this classifier. The classification of new classes is based on similarity measures such as distance matrix, Euclidean distance is mostly used in distance matrix of this classifier. KNN

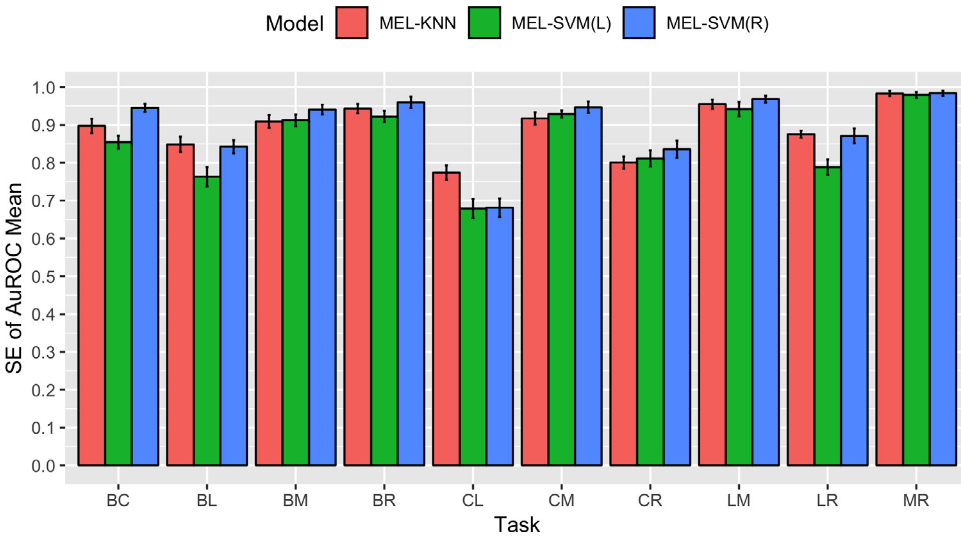

Fig. 15. Classification AUROC of subject 1 for 10 different combination of tasks.

Table 2
Accuracy of classification of 2-class using different classifiers.

Subject	Classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	89.5	85	91.25	94.25	76.75	91.5	79.75	95.25	87.25	98.25
	MEL-SVM(L)	85.5	75.75	90.75	92.25	67.75	92.75	81	93.75	79.5	98
	MEL-SVM(R)	94.5	84	93.75	96	67.5	95	83.75	96.75	87.75	98.5
Sub2	MEL-KNN	84	85	88	99	73	75	96	82.5	98	93
	MEL-SVM(L)	88.5	79.5	90.5	99.5	66.5	79.5	97.5	80.5	99	95.5
	MEL-SVM(R)	89.5	87.5	93	99.5	74	89.5	98	93	99	95
Sub3	MEL-KNN	48.25	64.75	79	60.25	72.5	78.75	61.75	86	69	85.25
	MEL-SVM(L)	55.25	68.75	86.75	61.5	70.75	78	56.5	86.5	66.75	78.75
	MEL-SVM(R)	64.25	69.75	87.75	64.5	72.75	81.5	64.5	86.5	71	83.75
Sub5	MEL-KNN	82	86.67	85.83	86.83	70.5	73.83	83.17	82	86	81.33
	MEL-SVM(L)	79.83	86.17	84.33	86.83	74.33	68.67	76	76.83	89.17	76.33
	MEL-SVM(R)	82.17	87.5	86.17	87	73.67	74.67	77.83	77.5	92.33	80.67
Sub6	MEL-KNN	78.75	65.25	84.5	84.5	71.25	88.5	81.75	77	80.25	77.75
	MEL-SVM(L)	82.25	63.25	87	82	73	95	82.5	81.25	75.25	80.75
	MEL-SVM(R)	86	66	88.75	86	76	96	85.25	90.75	79.75	84.25
Sub7	MEL-KNN	88.5	82.5	96.5	100	82	83	99	80.5	99	86.5
	MEL-SVM(L)	92.5	88	94.5	99.5	77	82.5	99	91.5	99.5	87.5
	MEL-SVM(R)	88	86.5	96	100	84.5	97.5	99.5	98	99.5	87

Table 3
Sensitivity of classification of 2-class using different classifiers.

Subject	classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	0.8824	0.8486	0.9449	0.9375	0.8971	0.9495	0.9005	0.9866	0.9123	0.9751
	MEL-SVM(L)	0.937	0.8846	0.9716	0.9575	0.8367	0.9603	0.9156	1	0.894	0.989
	MEL-SVM(R)	0.9311	0.9475	0.9857	0.9749	0.9186	0.9685	0.891	1	0.8498	0.983
Sub2	MEL-KNN	1	0.9723	0.9846	1	0.8906	0.8895	0.9155	0.9044	0.9608	0.8708
	MEL-SVM(L)	0.9715	0.9295	0.984	1	0.8447	0.9513	1	0.926	0.9933	0.9693
	MEL-SVM(R)	0.9826	0.9692	0.9279	1	0.8911	0.89	0.9917	0.9442	0.9933	0.951
Sub3	MEL-KNN	0.7774	0.9343	0.9684	0.8433	0.9044	0.9395	0.7822	0.9299	0.8279	0.9293
	MEL-SVM(L)	0.7878	0.8803	0.9645	0.8255	0.8871	0.9481	0.7029	0.9293	0.7947	0.9082
	MEL-SVM(R)	0.8369	0.8367	0.945	0.8263	0.8877	0.9097	0.7898	0.9068	0.8268	0.941
Sub5	MEL-KNN	0.8987	0.9325	0.8898	0.8604	0.8706	0.7391	0.8134	0.873	0.854	0.8515
	MEL-SVM(L)	0.9101	0.9623	0.9594	0.9575	0.8734	0.8231	0.8481	0.8817	0.9453	0.8675
	MEL-SVM(R)	0.9138	0.9599	0.9353	0.9313	0.8624	0.8599	0.8382	0.8848	0.9458	0.8867
Sub6	MEL-KNN	0.8689	0.6311	0.8459	0.8498	0.6397	0.9445	0.8987	0.9302	0.9243	0.9344
	MEL-SVM(L)	0.9258	0.7639	0.9477	0.9101	0.8484	0.9896	0.9216	0.9233	0.8796	0.8716
	MEL-SVM(R)	0.9311	0.6753	0.957	0.9075	0.8713	0.9938	0.9127	0.9151	0.8955	0.9209
Sub7	MEL-KNN	0.9826	0.8627	0.9917	1	0.8104	0.9622	0.9917	1	1	0.9833
	MEL-SVM(L)	0.9909	0.9357	0.9889	1	0.9332	0.9575	0.99	0.9909	1	0.9644
	MEL-SVM(R)	1	1	1	1	0.8905	0.97	0.9909	0.9798	1	0.9818

Table 4
Specificity of classification of 2-class using different classifiers.

Subject	classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	0.9123	0.8486	0.874	0.9486	0.652	0.8849	0.7009	0.9238	0.8383	0.9917
	MEL-SVM(L)	0.7722	0.6423	0.8531	0.8875	0.5216	0.8976	0.7079	0.8841	0.6836	0.9699
	MEL-SVM(R)	0.9595	0.7374	0.8963	0.9452	0.4438	0.9252	0.7809	0.9375	0.8926	0.9855
Sub2	MEL-KNN	0.7097	0.7228	0.7814	0.9812	0.5718	0.6248	1	0.7736	0.9923	0.9875
	MEL-SVM(L)	0.8041	0.6607	0.8476	0.9917	0.4753	0.6533	0.9499	0.696	0.9909	0.9399
	MEL-SVM(R)	0.8124	0.7691	0.9383	0.9889	0.5841	0.9076	0.9715	0.9212	0.9857	0.9513
Sub3	MEL-KNN	0.194	0.362	0.6104	0.3665	0.5475	0.632	0.4623	0.7892	0.5581	0.7902
	MEL-SVM(L)	0.3544	0.5125	0.7705	0.407	0.5424	0.6084	0.4282	0.803	0.5355	0.6647
	MEL-SVM(R)	0.5029	0.5665	0.8123	0.4646	0.5776	0.7158	0.5017	0.8295	0.601	0.7317
Sub5	MEL-KNN	0.7437	0.7991	0.8273	0.877	0.5383	0.7375	0.8512	0.7665	0.8664	0.7798
	MEL-SVM(L)	0.6874	0.7599	0.7243	0.7769	0.6173	0.5475	0.6726	0.6532	0.8381	0.6598
	MEL-SVM(R)	0.7396	0.7893	0.7881	0.8091	0.6141	0.6337	0.7202	0.668	0.9009	0.727
Sub6	MEL-KNN	0.7159	0.6873	0.8509	0.8485	0.7893	0.8301	0.7413	0.6224	0.6867	0.6249
	MEL-SVM(L)	0.7237	0.4902	0.7878	0.7227	0.5961	0.9139	0.7228	0.699	0.6141	0.747
	MEL-SVM(R)	0.7937	0.6427	0.8081	0.8032	0.646	0.9328	0.7954	0.9005	0.6805	0.767
Sub7	MEL-KNN	0.8133	0.8031	0.9526	1	0.8346	0.7048	0.9909	0.6221	0.9832	0.7553
	MEL-SVM(L)	0.8694	0.8169	0.8981	0.9917	0.6119	0.6615	0.9917	0.8434	0.9917	0.789
	MEL-SVM(R)	0.7719	0.7265	0.9148	1	0.7949	0.9775	1	0.9764	0.9889	0.7588

Table 5
MCC of classification of 2-class using different classifiers.

Subject	classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	0.7909	0.7006	0.8252	0.8875	0.568	0.8363	0.6078	0.9095	0.7486	0.9662
	MEL-SVM(L)	0.7171	0.5415	0.8267	0.847	0.3749	0.8554	0.6323	0.888	0.5946	0.961
	MEL-SVM(R)	0.8927	0.6962	0.8818	0.9202	0.4077	0.8991	0.6746	0.9373	0.7474	0.9695
Sub2	MEL-KNN	0.7366	0.7145	0.7749	0.9801	0.4807	0.523	0.9212	0.6768	0.9584	0.8657
	MEL-SVM(L)	0.7852	0.6143	0.8259	0.9903	0.3511	0.628	0.9482	0.6401	0.9786	0.9125
	MEL-SVM(R)	0.8018	0.7514	0.8598	0.9903	0.5011	0.7942	0.9613	0.8653	0.9774	0.9014
Sub3	MEL-KNN	−0.0398	0.3542	0.6217	0.2378	0.4882	0.6002	0.2613	0.7289	0.4005	0.7251
	MEL-SVM(L)	0.169	0.4221	0.7463	0.2533	0.4518	0.5869	0.1354	0.738	0.3458	0.5911
	MEL-SVM(R)	0.3731	0.4199	0.7608	0.3079	0.4906	0.6376	0.3031	0.7344	0.4382	0.6881
Sub5	MEL-KNN	0.6504	0.7424	0.7196	0.7378	0.4329	0.478	0.6667	0.6519	0.7211	0.635
	MEL-SVM(L)	0.6141	0.7392	0.7056	0.7465	0.5066	0.3872	0.5303	0.5538	0.7903	0.5411
	MEL-SVM(R)	0.6615	0.7646	0.7345	0.747	0.4921	0.5078	0.5632	0.5677	0.8488	0.6245
Sub6	MEL-KNN	0.5921	0.3219	0.696	0.7025	0.4324	0.7779	0.6454	0.582	0.6343	0.5872
	MEL-SVM(L)	0.6618	0.2585	0.7464	0.645	0.4633	0.9046	0.6569	0.6371	0.5092	0.6193
	MEL-SVM(R)	0.7343	0.3169	0.7819	0.7158	0.535	0.9235	0.7156	0.8127	0.5891	0.695
Sub7	MEL-KNN	0.799	0.6663	0.9343	1	0.6478	0.6832	0.9807	0.6675	0.9803	0.7523
	MEL-SVM(L)	0.8619	0.7686	0.8893	0.9903	0.578	0.6527	0.9807	0.8405	0.9903	0.763
	MEL-SVM(R)	0.7883	0.7599	0.9184	1	0.6975	0.9511	0.9905	0.9609	0.9903	0.7593

Table 6
Kappa values for classification of 2-class using different classifiers.

Subject	classifier	BC	BL	BM	BR	CL	CM	CR	LM	LR	MR
Sub1	MEL-KNN	0.783	0.6959	0.8215	0.8837	0.5409	0.8292	0.5901	0.9048	0.7411	0.9646
	MEL-SVM(L)	0.6811	0.5041	0.7807	0.8332	0.3688	0.8575	0.6085	0.8835	0.5811	0.9644
	MEL-SVM(R)	0.8688	0.6268	0.8669	0.9088	0.4047	0.8884	0.7144	0.9291	0.7605	0.9848
Sub2	MEL-KNN	0.6975	0.688	0.7527	0.9792	0.4548	0.4933	0.915	0.6518	0.9562	0.8562
	MEL-SVM(L)	0.7172	0.5999	0.7935	0.99	0.4043	0.5113	0.9268	0.6468	0.9585	0.9379
	MEL-SVM(R)	0.7601	0.7114	0.846	0.99	0.477	0.7329	0.9786	0.8465	0.9894	0.8872
Sub3	MEL-KNN	−0.0285	0.2948	0.5776	0.2049	0.4503	0.57	0.24	0.717	0.3812	0.7077
	MEL-SVM(L)	0.1247	0.3864	0.7162	0.2143	0.4707	0.5681	0.1334	0.7242	0.3473	0.6171
	MEL-SVM(R)	0.2795	0.4318	0.7814	0.3677	0.4903	0.6288	0.24	0.7267	0.4448	0.6869
Sub5	MEL-KNN	0.6401	0.7325	0.7164	0.7365	0.4075	0.4758	0.6632	0.6392	0.7196	0.6277
	MEL-SVM(L)	0.6141	0.7268	0.6228	0.7133	0.4848	0.3871	0.5289	0.5191	0.7896	0.5337
	MEL-SVM(R)	0.6526	0.7836	0.764	0.7465	0.5117	0.4302	0.5471	0.5661	0.8195	0.5872
Sub6	MEL-KNN	0.5772	0.3111	0.6889	0.6912	0.4239	0.7699	0.6336	0.549	0.6073	0.5562
	MEL-SVM(L)	0.589	0.2381	0.7332	0.6183	0.4737	0.8885	0.6141	0.6454	0.5028	0.5913
	MEL-SVM(R)	0.6705	0.3633	0.7903	0.7317	0.504	0.8787	0.7163	0.763	0.6469	0.6672
Sub7	MEL-KNN	0.7744	0.6534	0.9281	1	0.6334	0.6549	0.9798	0.6131	0.9794	0.7262
	MEL-SVM(L)	0.8604	0.7221	0.8981	1	0.6339	0.6806	0.9898	0.8554	0.9886	0.7137
	MEL-SVM(R)	0.7813	0.6885	0.8988	1	0.5978	0.9596	0.9898	0.97	0.9794	0.7276

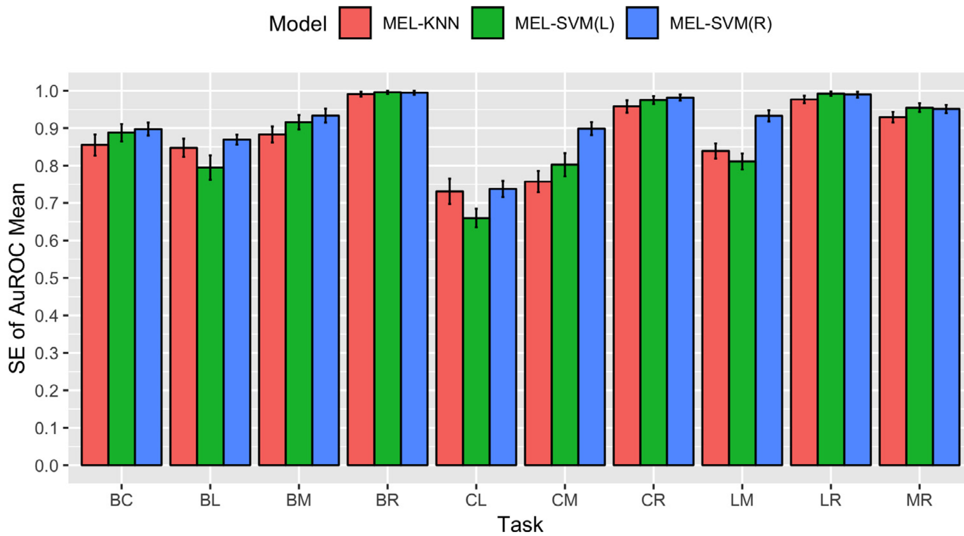

Fig. 16. Classification AUROC of subject 2 for 10 different combination of tasks.

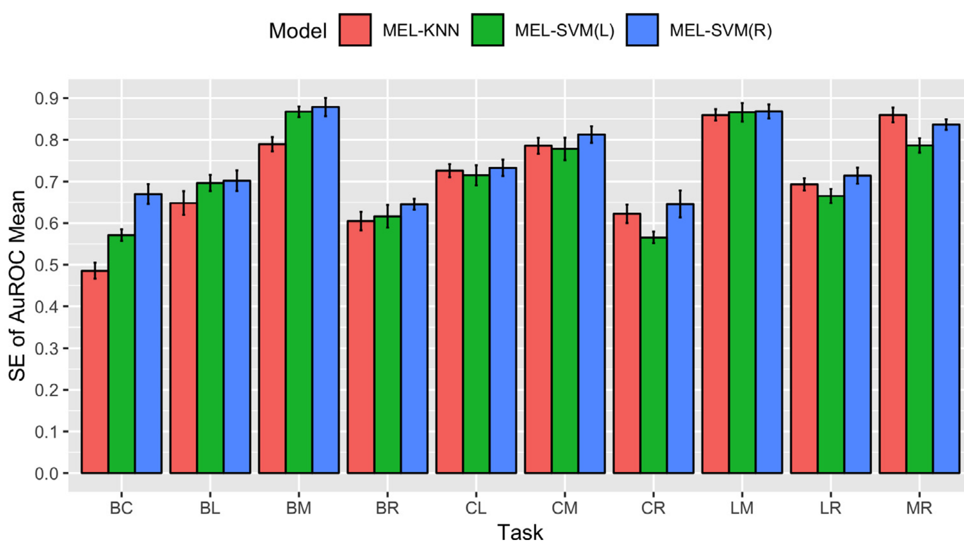

Fig. 17. Classification AUROC of subject 3 for 10 different combination of tasks.

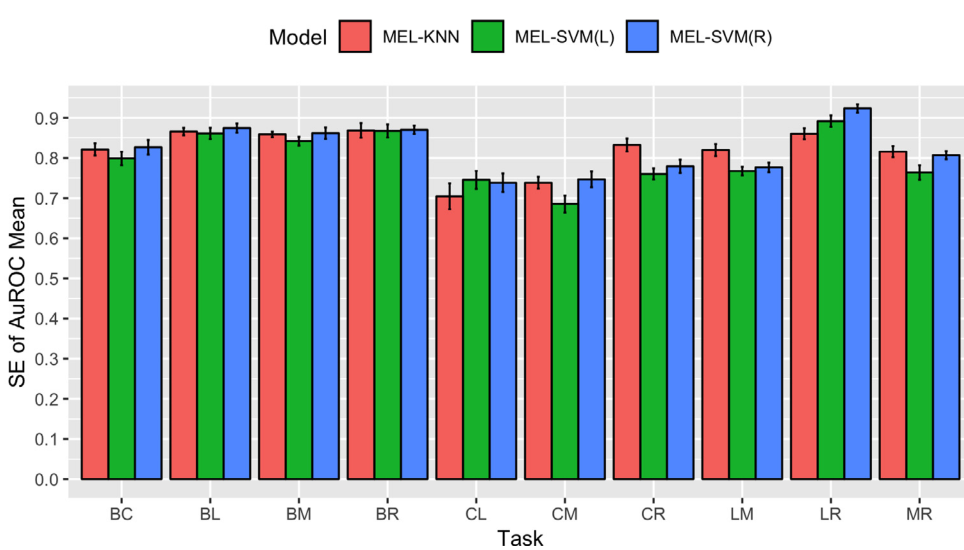

Fig. 18. Classification AUROC of subject 5 for 10 different combination of tasks.

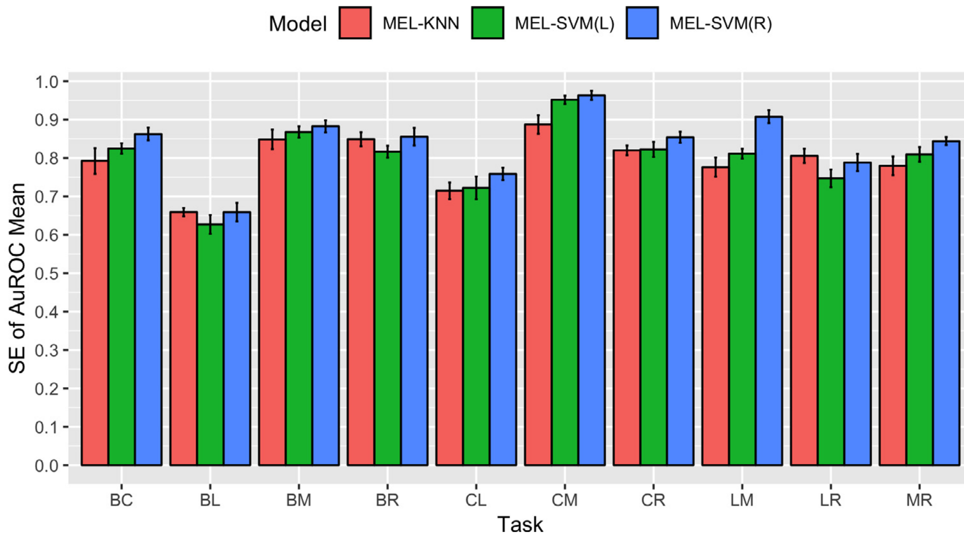

Fig. 19. Classification AUROC of subject 6 for 10 different combination of tasks.

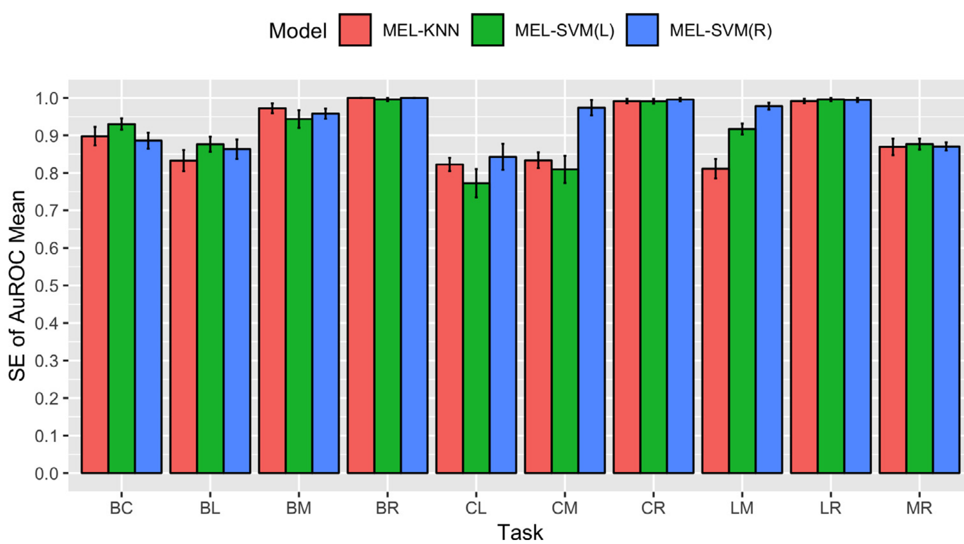

Fig. 20. Classification AUROC of subject 7 for 10 different combination of tasks.

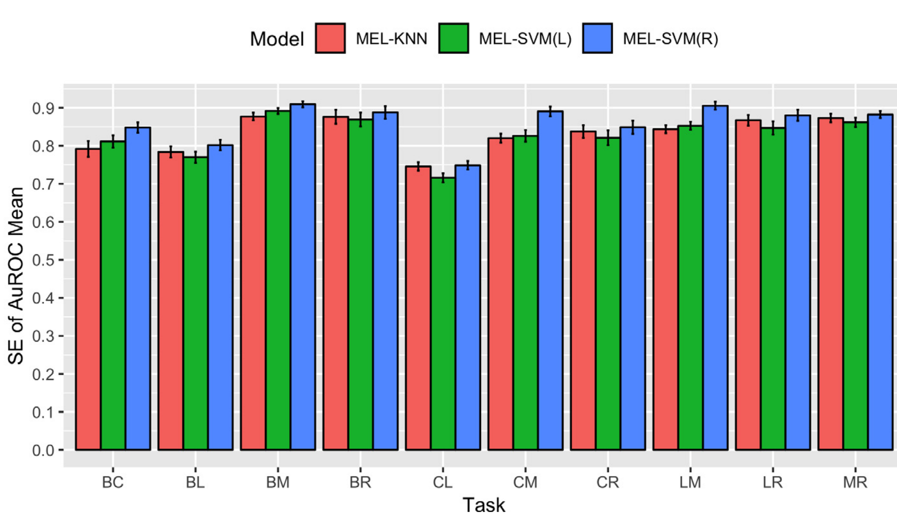

Fig. 21. Average classification AUROC of all subjects for 10 different combination of tasks.

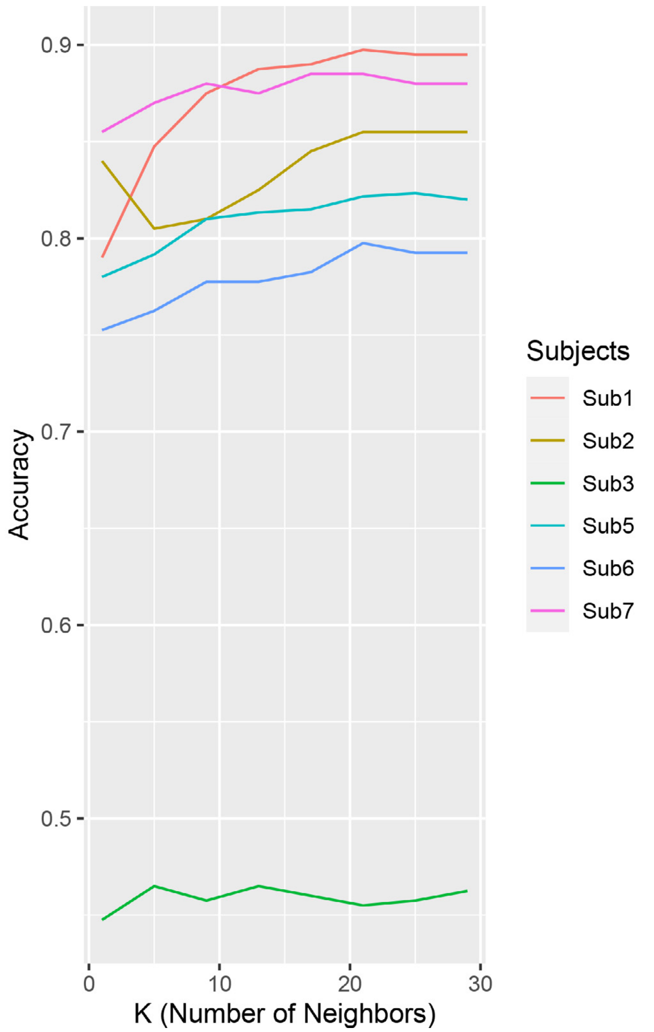

Fig. 22. Variation in the accuracy of all subjects for the BC task as the number of neighbors(k)(k) changed.

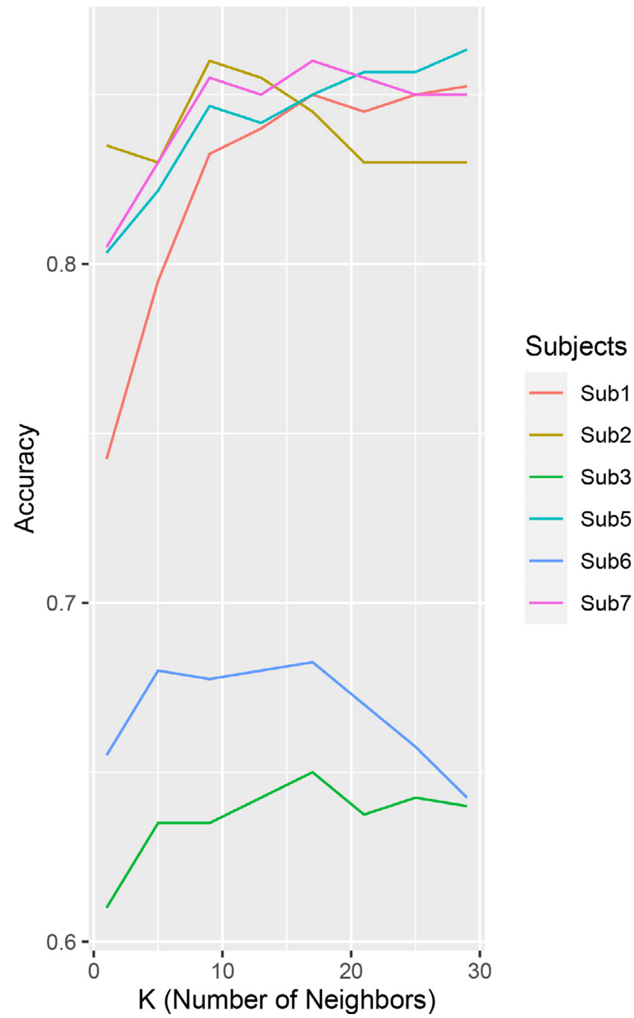

Fig. 23. Variation in the accuracy of all subjects for the BL task as the number of neighbors(k) changed.

has a drawback if there is large training data because large time is required to find the nearest neighbor [70]. To overcome this drawback, dimension reduction techniques are used. The following procedure is used to find the class of x in KNN classifier:

1. Based on the distance measure determine the k instance to the nearest class x.
2. Now the k instant is allowed to vote for finding the class x.

The results obtained are used for calculating the performance using different measures.

5.4. Classification performance and evaluation

Classification performance on the different binary combinations of the mental tasks on given subjects is evaluated using the area under the receiver operating characteristics curve (AUROC) [71]. Where the AUROC score of up to 0.5 is a random guess, and a score of 1.0 shows perfect prediction by the model. Other useful performance parameters like accuracy (acc) [72], sensitivity (sn) [72], specificity (sp) [72], and Matthews correlation coefficient (MCC) have also been reported along with the mean and standard deviation on 10-fold test examples. These parameters were computed from the confusion matrix of prediction results as follows:

$$acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100 \quad \in [0, 100], \tag{27}$$

$$sn = \frac{TP}{TP + FN} \quad \in [0, 1], \tag{28}$$

$$sp = \frac{TN}{FP + TN} \quad \in [0, 1], \tag{29}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \in [-1, +1]. \tag{30}$$

where TP, TN, FP, and FN represent the number of positively labeled examples predicted as positive, the number of negatively labeled examples predicted as negative, the number of negatively labeled examples predicted as positive and the number of positively labeled examples predicted as negative, respectively.

In binary classification, the class prediction for each instance is often made based on a X, which is a "score" computed for the instance (e.g. the estimated probability in logistic regression). Given a threshold parameter T, the instance is classified as "positive" if $X > T$, and "negative" otherwise. X follows a probabil-

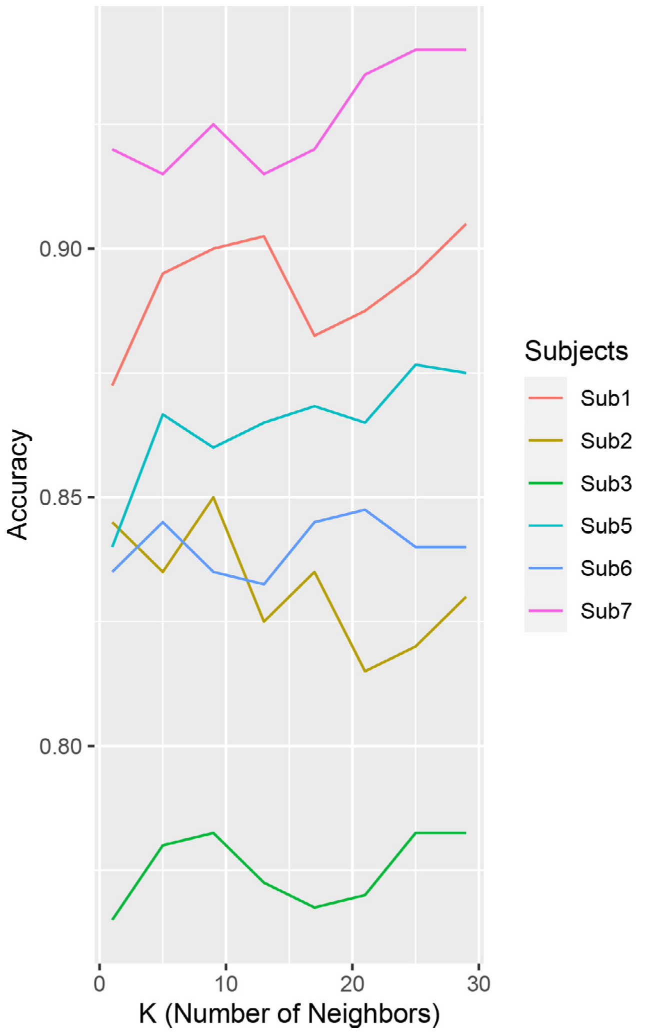

Fig. 24. Variation in the accuracy of all subjects for the BM task as the number of neighbors(k) changed.

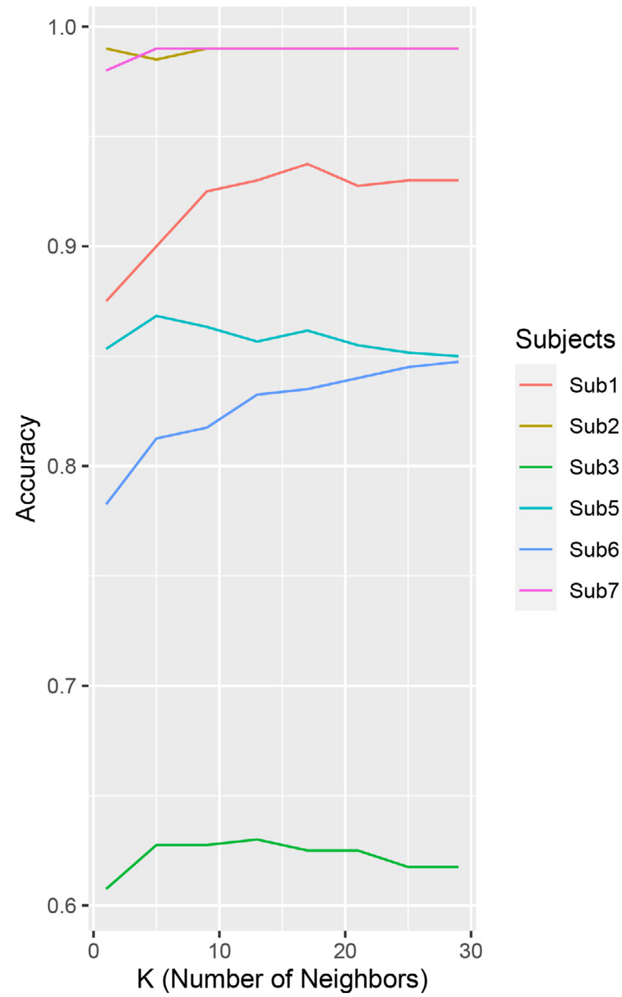

Fig. 25. Variation in the accuracy of all subjects for the BR task as the number of neighbors(k) changed.

ity density $f_1(x)$ if the instance actually belongs to class "positive", $f_0(x)$ if otherwise. Therefore, the true positive rate is given by $\text{TPR}(T) = \int_T^\infty f_1(x)\,dx$ and the false positive rate is given by $\text{FPR}(T) = \int_T^\infty f_0(x)\,dx$. The ROC curve plots parametrically TPR(T) versus FPR(T) with T as the varying parameter [71].

$$\begin{aligned}
TPR &= \tfrac{TP}{TP+FN}, \\
TNR &= \tfrac{TN}{TN+FP}, \\
AUROC &= \int_{x=0}^{1} \text{TPR}\left(\text{FPR}^{-1}(x)\right)dx, \\
&= \int_{\infty}^{-\infty} \text{TPR}(T)\text{FPR}'(T)dT, \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(T' > T)f_1(T')f_0(T)dT'dT, \\
&= P(X_1 > X_0).
\end{aligned} \tag{31}$$

For evaluation and comparison of different classifiers, we used k-fold cross validation algorithm. In this method, datasets are segmented into two parts; one for training and other for testing the classification algorithm. Firstly, the dataset is segmented into k equally sized data sets. Total k number of iteration are performed for training and testing so that for every iteration $k-1$ different sets of segmented data are used for training and the remaining set is held out for validation. To get rid of overlapping between validation and training set, which can lead to poor performance of the classifier, to avoid this k-fold cross validation is used. Therefore, the accurate performance of classification algorithm can be obtained using k-fold cross validation. In our analysis, we have chosen $k = 10$, which utilizes 90% data for prediction (Table 1).

6. Results and discussions

6.1. Dataset

In our experiment, we have utilized freely accessible data for mental task classification [73]. Seven subjects were taken in original EEG data; we used information from all subjects except the subject 4 because of some missing data. Each subject performed five distinctive mental assignments: the baseline task (B)(no task), the mental letter composing task (L), the nontrivial mathematical task (M), the visualizing counting of numbers written on a blackboard task (C), and the geometric figure rotation task (R). Each session consists of five trials and each trial has five mental tasks. EEG signal recording was done by placing six electrodes on the scalp at C3, C4, P3, P4, O1, and O2 referencing to two electrodes put at electrically connected mastoid, A1, and A2, as appeared in Fig. 4. Every trial is recorded with sampling frequency of 250 Hz and have total duration of 10 s, per trial consists of 2500 samples points. For more information about data can be found in the work of Keirn and Aunon [73]. Following are the task defined in the dataset:

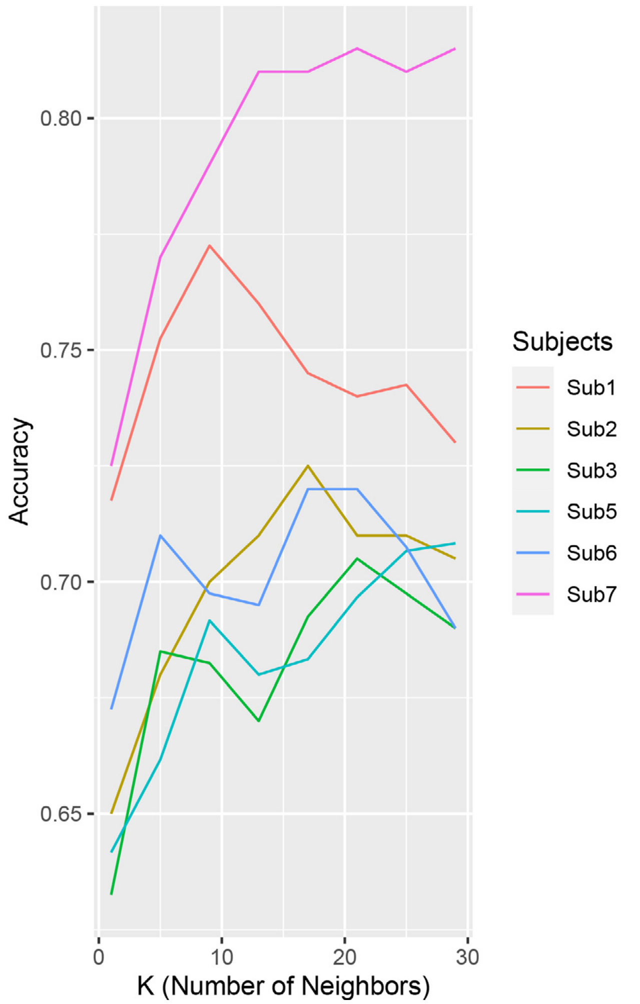

Fig. 26. Variation in the accuracy of all subjects for the CL task as the number of neighbors(k) changed.

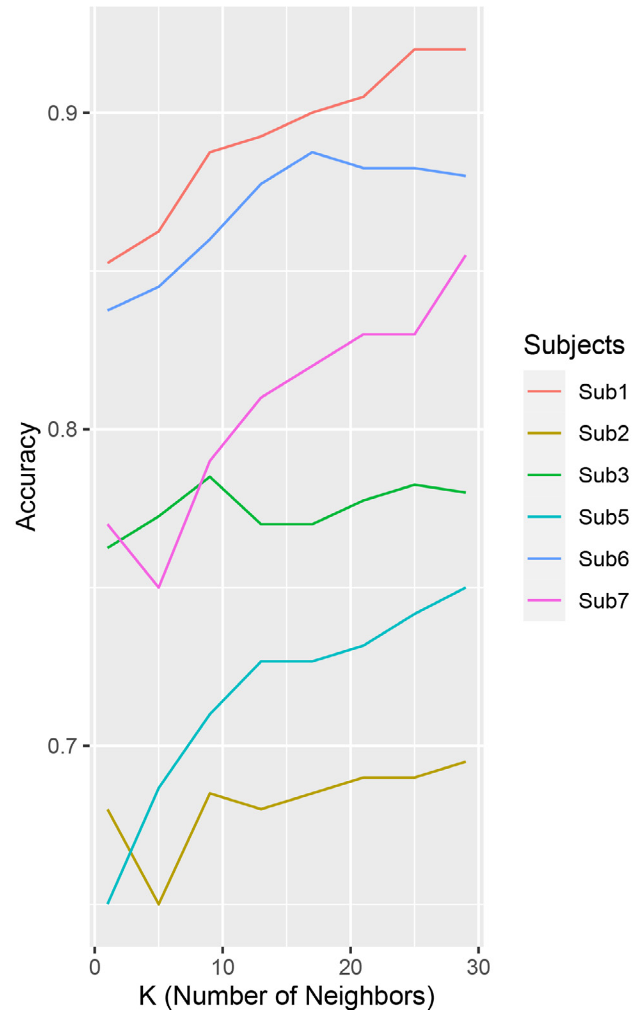

Fig. 27. Variation in the accuracy of all subjects for the CM task as the number of neighbors(k) changed.

1. Baseline task: subjects were advised to remain relax.
2. Mathemetic task: subjects were advised to solve a non-trivial mathematical problem for example 49 times 78.
3. Letter task: subjects were advised to mentally think for a letter.
4. Rotation task: subjects were advise to think about a 3-dimensional block which is rotating about an axis.
5. Counting task: subjects were advise to think about a black-board in which numbers are written sequentially.

6.2. Subject 1

Among the subjects when pair wise classification of task was performed then highest AUROC for baseline vs count and multiplication vs rotation were obtained for subject 1 as shown in Fig. 5 and Fig. 6–8 when MEL SVM with non linear kernel was used. The highest AUROC was obtained for the letter-math (LM) and math-rotation (MR) combination. This high AUROC might be due the fact of maths task was performed by the right hemisphere while rotation or letter task was performed by left hemisphere. Therefore, the combination of LM or MR tasks gives better results. Fig. 15 clearly shows, among three classifier, the SVM with non linear kernel performance better than the MEL-KNN and MEL-SVM with linear kernel. The lowest AUROC is yield by count-letter (CL) combination, which is lower than 90% for subject 1. The average AUROC of 0.9834 for subject 1 is obtained when MEL-KNN classifier is used with 10-fold cross validation method. Highest average AUROC of 0.9850 is obtained when MEL-SVM with non linear kernel is used with 10-fold cross validation method, which is the highest AUROC obtained for subject 1 using non linear kernel. Similarly other parameters of performance such as acc, sn, sp, and MCC for subject 1 are given in Table 2, Table 3, Table 4, and Table 5 respectively. As shown in Table 6, kappa values for all the classification tasks except BL and CL are higher for subject 1 when MEL-SVM(R) was used and the highest kappa value of 0.9848 was obtained for MR combination by using MEL-SVM with the non linear kernel.

6.3. Subject 2

The combination of letter-rotation (LR) and baseline-rotation (BR) obtained highest AUROC for subject 2. In Fig. 16, the highest average AUROC of 0.9958 for subject 2 was obtained using 10-fold cross validation method, and MEL-SVM classifier with linear kernel. The highest accuracy of 98.7% as shown in Table 2 was obtained for MR combination when MEL-SVM with non linear kernel was used. Sn of 1 for the combination of BC, BR, and CR was achieved for subject 2. Sp for CR combination was 1 which is highest among all the combination. MCC for subject 2 was highest for BR combination as compare to other combination which is shown in Table 5. As shown in Table 6, kappa values for all the classification tasks except BR and MR are higher for subject 2 when MEL-SVM(R) was used and the highest kappa value of 0.99 was

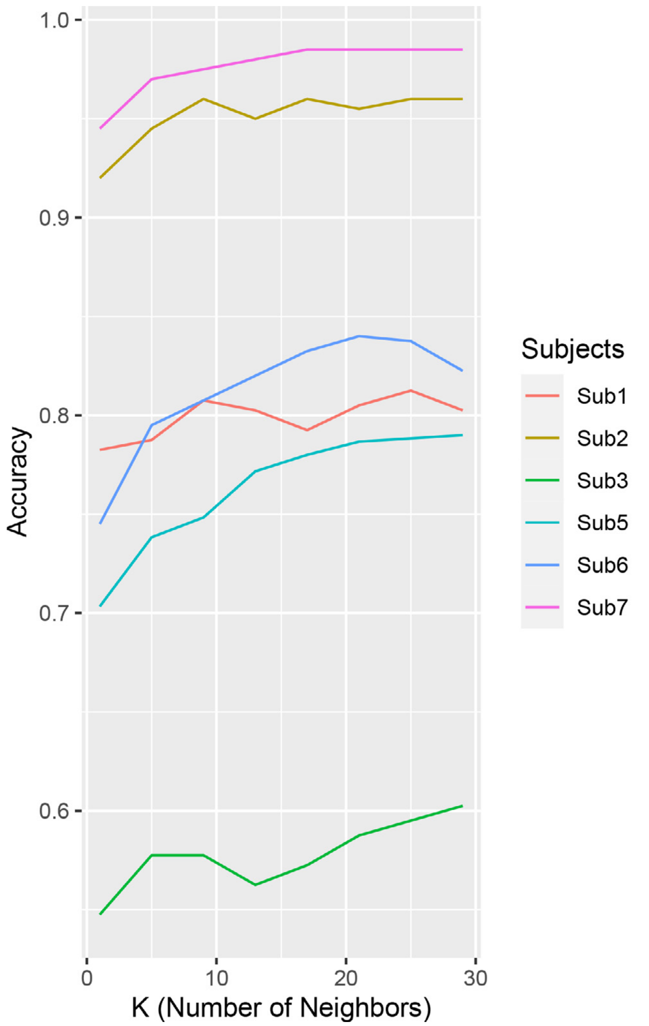

Fig. 28. Variation in the accuracy of all subjects for the CR task as the number of neighbors(k) changed.

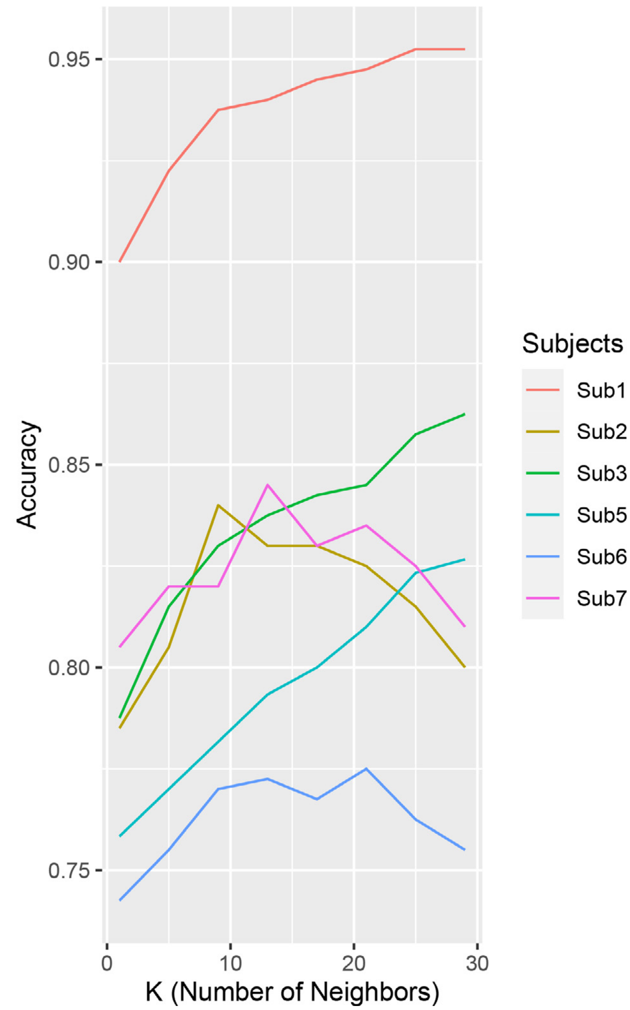

Fig. 29. Variation in the accuracy of all subjects for the LM task as the number of neighbors(k) changed.

obtained for BR combination by using MEL-SVM with the linear and non linear kernels.

6.4. Subject 3

Fig. 17 shows the classification results for different task combinations using different classifiers, BM, LM and MR combination achieved highest AUROC. For the validation of our results we have used 10-fold cross validation. The highest acc of 87.75% was obtained for BM as shown in Table 2, when MEL-SVM with non linear kernel was used. The highest sn, sp, and MCC were achieved for BM, LM, and BM combination as shown in Table 3, Table 4, and Table 5 respectively. The average AUROC for LM and MR combination using MEL-KNN are given as 0.8596 and 0.8597 respectively. The average AUROC for BM and LM combination using MEL-SVM with linear is 0.8675 and 0.8662, and with non linear kernels is 0.8786 and 0.8682 respectively. For subject 3 as shown in Table 6, the highest kappa values for all combination tasks except CR were obtained when MEL-SVM with the non linear kernel was used. The highest kappa value 0.7814 was obtained for the BM task.

6.5. Subject 5

Table 2 shows the acc for different combination of tasks in which the highest acc of 92.33% for subject 5 is achieved for LR combination when MEL-SVM with non linear kernel is used. The highest sn as shown in Table 3 of 0.9623 is achieved for BL combination. When MEL-KNN classifier was used the highest AUROC is obtained for BR and LR combination. MEL-SVM with linear kernel gives better performance for the combination of BR and LR. MEL-SVM with non linear kernel gives better performance for the combination of baseline-letter and letter-rotation combination.The average AUROC for BR and LR combination using MEL-KNN are given as 0.8687 and 0.8602 respectively. The average AUROC for BR and LR combination using MEL-SVM with linear is 0.8672 and 0.8917, and the average AUROC for BL and LR combination using MEL-SVM with non linear is 0.8746 and 0.9233, using 10-fold cross validation method. In Table 4, it can be seen that the highest sp of 0.9009 is obtained for LR combination and MCC of 0.8488 for same combination which is given in Table 5. Both the sp and MCC are obtained by using MEL-SVM with non linear kernel. For the subject 5 MEL-KNN have better kappa values for CM, CR, LM, and MR as shown in Table 6, while for the other classification tasks kappa values are better for when MEL-SVM with the non linear kernel was used. The highest kappa value of 0.8195 was obtained for LR combination when MEL-SVM with the non linear kernel was used.

6.6. Subject 6

By using MEL-SVM with non linear kernel, the highest acc of 96% is achieved for CM combination for subject 2 as shown in

Fig. 30. Variation in the accuracy of all subjects for the LR task as the number of neighbors(k) changed.

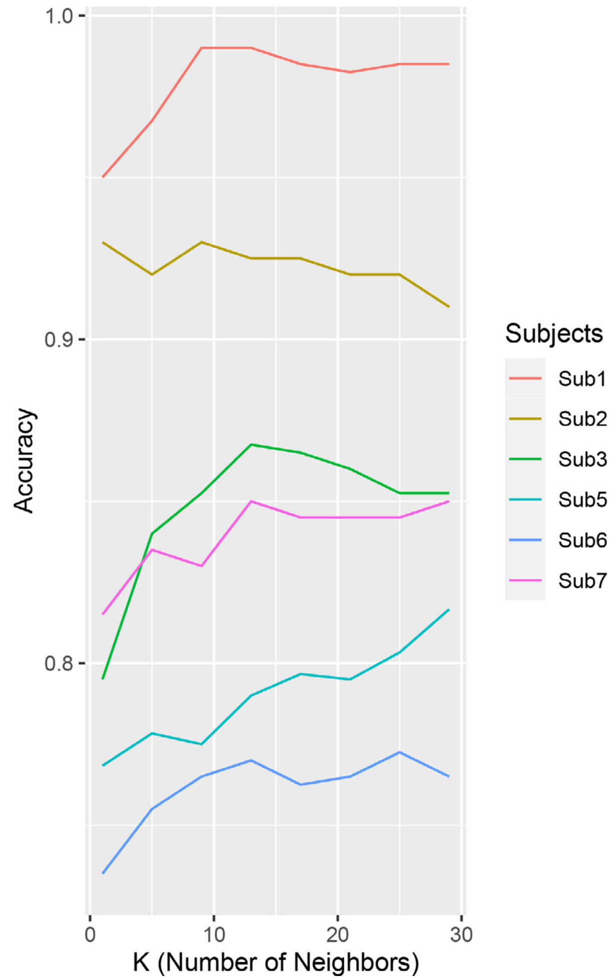

Fig. 31. Variation in the accuracy of all subjects for the MR task as the number of neighbors(k) changed.

Table 2. In Table 3, it can be seen that, the sn of 0.9938 by using MEL-SVM with non linear kernel for CM combination is achieved. The BM, BR, CM and LM combination tasks obtained, the highest AUROC as shown in Fig. 19 by using different classifiers. The average AUROC for BR and CM combination using MEL-KNN are given as 0.8491 and 0.8873 respectively. The average AUROC for BM and CM combination using MEL-SVM with linear is 0.8677 and 0.9517, and the average AUROC for CM and LM combination using MEL-SVM with non linear is 0.9633 and 0.9078, using 10-fold cross validation method. The parameters MCC and sp are obtained by using MEL-SVM with non linear kernel. In Table 5, it can be seen that the highest MCC of 0.9235 is obtained for CM combination and sp of 0.9328 for same combination which is given in Table 4. The kappa values for different combination of tasks for subject 6 is shown in Table 6 in which highest kappa values were obtained for all the combination of tasks except CM when MEL-SVM with the non linear kernel was used. The highest kappa value of 0.8885 was obtained for CM task when MEL-SVM with the linear kernel was used.

6.7. Subject 7

Among the subjects when pair wise classification of task was performed then highest AUROC except baseline vs count and multiplication vs rotation were obtained for subject 7 as shown from Fig. 6 to Fig. 13. For BM, BR, and CL MEL KNN classier gives the highest AUROC. For BL and LR, MEL SVM with linear kernel achieved highest AUROC and for other combinations MEL SVM with non linear kernel gives highest AUROC for subject 7 which is higher than other subjects. When MEL-SVM with linear and non linear kernel was used then highest acc of 100% was achieved for the BL combination as shown in Table 2. The sn of 1 is achieved for all the three MEL classifier for the combination of BR and LR, which is depicted in Table 3. In Fig. 20, it is clearly shown that the highest AUROC was obtained for the BR, CR and LR combination by using different classifiers. The average AUROC for BM and CM combination using MEL-SVM with linear is 0.9958 and 0.9958, and the average AUROC for CM and LM combination using MEL-SVM with non linear is 1.00 and 0.9955, also the average AUROC for BR and CM combination using MEL-KNN are given as 1.00 and 0.9916 respectively. All these parameters were validated using 10-fold cross validation method. The performance parameter sp and MCC were also evaluated for different combinations. The highest sp as given in Table 4 of 1 was obtained for combination BR and CR. Table 5 shows the MCC of all subjects, for subject 7, highest MCC of 1 was achieved for the combination of BR when MEL-KNN and MEL-SVM with non linear kernel was used. Table 6 shows the kappa values for different combination of tasks, in which for subject 7 the highest kappa values for classification tasks were obtained by MEL-SVM with the linear kernel followed by MEL-SVM with the non linear kernel and the highest kappa value of 1 was obtained for the BR task by all the three classifiers.

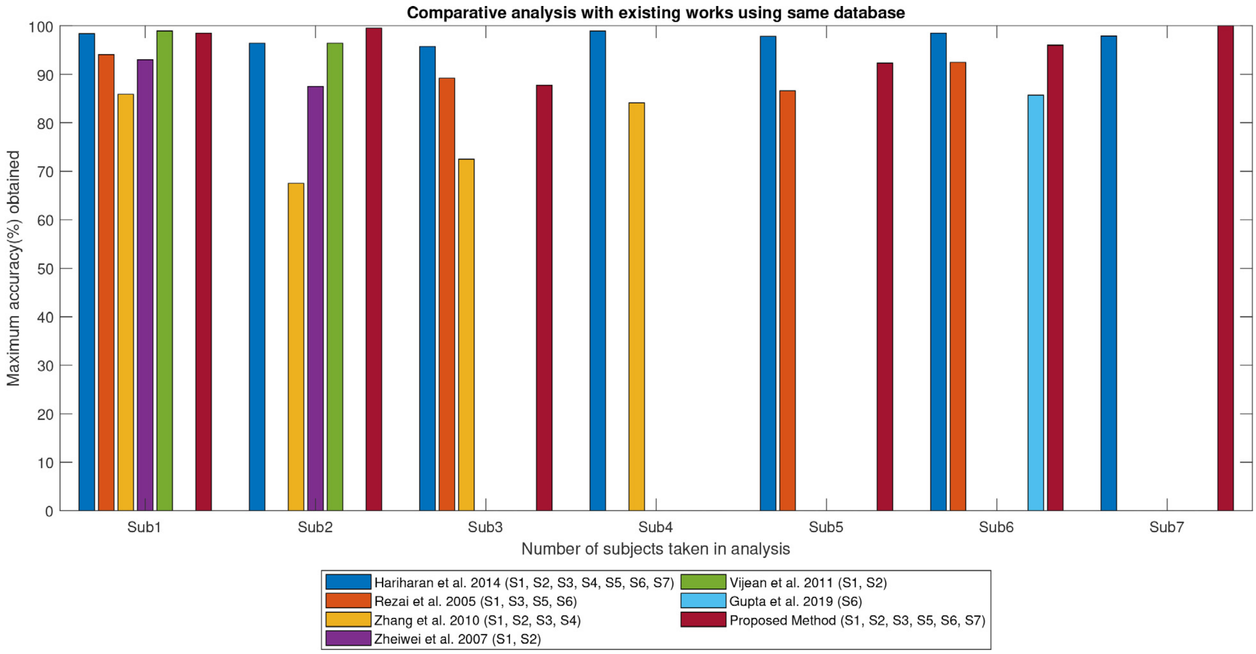

Fig. 32. Comparative analysis with existing works using same database.

6.8. All Subjects

Among the subjects when pair wise classification of task was performed then highest AUROC for baseline vs count and multiplication vs rotation were obtained for subject 1 as shown in Fig. 5 and Fig. 7 when MEL SVM with non linear kernel was used. For subject 7 the highest AUROC were obtained for the classification task BM, BR, and CL as shown in Fig. 6, Fig. 8, and Fig. 9 respectively when MEL KNN classier was used. For BL and LR as shown in Fig. 6, and Fig. 13 MEL SVM with linear kernel achieved highest AUROC and for combinations as shown in Fig. 9, Fig. 10, Fig. 11, and Fig. 12 were achived when MEL SVM with non linear kernel was used. The Fig. 15–19 represents average AUROC of all classifiers for all the 10 task combinations. The highest average AUROC is obtained for the subject 7 suggests that the MEL-KNN classifier and MEL-SVM with non linear kernel classifier are best suited for 2 class problem of mental task classification. Variation in AUROC is also observed for different combination of task using different classifier. For subject 1, the best AUROC is obtained for math-rotation task using MEL-SVM classifier with non linear kernel. For subject 2, the best AUROC is obtained for baseline-rotation task using MEL-SVM classifier with linear kernel. For subject 3, the best AUROC is obtained for baseline-math task using MEL-SVM classifier with non linear kernel. For subject 5, the best AUROC is obtained for letter-rotation task using MEL-SVM classifier with non linear kernel. For subject 6, the best AUROC is obtained for count-math task using MEL-SVM classifier with non linear kernel. For subject 7, the best AUROC is obtained for baseline-rotation task using MEL-KNN classifier and MEL-SVM classifier with non linear kernel.

The Table 2 shows the average accuracy including MEL based three classifier. In which highest accuracy is obtained by using MEL-SVM classifier with non linear kernel. Most of the classification task obtained accuracies which are higher than existing methods. So our proposed algorithm obtained improved results of average classification in terms of maximum classification as shown by existing experiment results on the same data set, for example Palaniappan and Huan [74] shows maximum accuracy of 92.7%, Anderson et al. [75] shows an overall accuracy of 91.4%. Similarly other parameters of performance such as acc, sn, sp, and MCC for all subjects are given in Table 3, Table 4, and Table 5 respectively. The kappa values for all the classification tasks for all the subjects is given in Table 6, in which generally highest kappa values for all the classification tasks were obtained by MEL-SVM with the non linear kernel except CL and BM. For BL task highest kappa value was obtained for subject 5 and for BM task it was obtained for subject 7. The highest kappa values of 1 were obtained by all the three classifiers for subject 7. It can be seen from Fig. 21, the highest value of average AUROC for all the subjects is obtained for the combination of tasks BM and LM when MEL-SVM with the non-linear kernel is used and the lowest for the combination of task CL is obtained when MEL-SVM with the linear kernel is used.

Effect of parameter on the study

In the study, the parameter k has been introduced at two places, first, to represent nearest neighbors count in KNN and second, to represent the number of views in the MEL model; however, four views have been used in the study. The effect of both of them on the model performance are discussed below:

Effect of k in KNN:

In k-nearest neighbor's algorithm (k-NN), a lower value of k may cause overfitting, whereas, the higher value of k smoothens the decision boundary and leads to an under-fitted model. Hence, the value of k is a critical parameter that determines the model performance. In this study, we used 10-fold cross-validation to get optimal value of k ranging from 1 to 30. The variation of accuracy for different values of k for all the subjects was shown from Figs. 22–31**?**.

Effect of k in MEL:

As we have discussed in the article, the MEL models make use of the consensus principle to label the test examples. The consensus aims to maximize the agreement on multiple distinct views. Therefore, we think the introduction of a very high number of views in the MEL model may over-smoothen the decision boundary, which can lead to the underfitting problem. Whereas, too low number of views may not reduce the error due to the overfitting problem. Hence, finding the optimal number of views for the classification of the given dataset is an optimization problem. However, in this study, we had only four views, and we think this number is neither too high nor too low for the study. In the future, if we can get features from more sources, then we can apply optimization techniques to find the optimal number of views for the given dataset.

6.9. Comparative analysis with existing methods

Fig. 32 shows the comparative analysis of the existing work on the same database with the proposed approach. From the figure, it can be seen that Rezaei et al. [76] have performed pairwise classification using autoregressive coefficients for mental tasks and obtained an accuracy of 94.07% for subject 1 while our proposed approach was able to give an accuracy of 98.5% for the subject 1. The obtained classification results for subject 3 5 6 by our proposed approach are higher. which outperforms the results given by Rezaei et al. in their study.

In [8], Zhang et al. also performed pairwise classification of mental task and obtained accuraies lower than our proposed approach in which they used lower frequency bands and power spectrum. From the Fig. 32, it can be inferred that Zhiwei et al. [77] also achieved classification accuracy for mental task in the range of 68.50% and 93.00 in which they used wavelet packet transform and SVM for classification, here also our proposed approach outperforms. Hariharan et al. [14] used S-transform based MSR as features and three classifiers KNN, LDA, and SVM for classification and achieved accuracies lower than the proposed approach in case of S1, S2, and S7 but higher in case of other subject for that we have used other classification performance measures such as sensitivity, specificity, and MCC which gives better performance results. Gupta et al. [10] used power spectral density and achieved an accuracy of 85.71% for only subject 6. Similarly other studies can also be outperform by our approach. Hence the proposed approach was able to out perform the existing work in literature. The analysis shows that, MEL approach can better for the classification of mental using EEG signals for BCI applications.

7. Conclusion and future work

Brain computer interface enables the physically challenged people to interact with the physical world using their brain EEG signal. WT and EMD can be successfully used to deal with non-linear and non-stationary nature of EEG to extract the relevant features of EEG signals. EWT along cannot deal with EEG, since EEG has a nature of overlapping in time and frequency domain. For the extraction of EEG signal's features FCM can be employed along with EWT, which eliminates the along application of EWT for EEG signal. Since the different features of signal give different results in different views. The features extracted by WT, EMD, FCM along with EWT and feature coding techniques can be combine using multiview learning. By using multiview learning, accuracy of classification can be enhanced. The results of multiview approach shows that there is variation in accuracies from subject to subject. The highest classification accuracy of 100% is obtained for the combination of baseline-rotation task. Future work will consider some other advance EEG feature extraction techniques and some more robust ensemble classifiers. These can improve the classification accuracy, which will enhance the interface between human brain and computer. Therefore this will lead to improvement in some application such as driving car using brain EEG signal and also prosthetic arm. In future our proposed method can be used to classify problems which involve multi-class mental tasks. It would be of more interest to perform a comparison of various existing classifiers in the literature [40–45] with SVM and KNN classifiers as a future work. In the future, advance EEG feature extraction techniques can be used to get more features, then we can apply optimization techniques to find the optimal number of views for the given dataset.

CRediT authorship contribution statement

A. Gupta: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing. **R.U. Khan:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing. **V.K. Singh:** Validation, Writing - review & editing, Resources, Software. **M. Tanveer:** Validation, Writing - review & editing, Resources, Software. **D. Kumar:** Validation, Writing - review & editing, Resources, Software. **A. Chakraborti:** Validation, Writing - review & editing, Resources, Software. **R.B. Pachori:** Validation, Writing - review & editing, Resources, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Department of Science & Technology (DST), Ministry of Science & Technology, Govt. of India under Cognitive Science Research Initiative (CSRI) Scheme, (Grant No. SR/CSI/PDF-05/2015 and Grant No. DST/ CSRI/PDF-50/2018), DST under Interdisciplinary Cyber Physical Systems (ICPS) - Data Science Research Initiative Scheme (Grant No. DST/ICPS/CPS-Individual/2018/276) and Council of Scientific & Industrial Research (CSIR), New Delhi, INDIA under Extra Mural Research (EMR) Scheme (Grant No. 22(0751)/17/EMR-II). We gratefully acknowledge Jawaharlal Nehru University (JNU) New Delhi and the Indian Institute of Technology Indore for providing facilities and support.

References

[1] K.S. Bayram, M.A. Kızrak, B. Bolat, Classification of EEG signals by using support vector machines, in: IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, 2013, pp. 1–3.
[2] D. Garrett, D.A. Peterson, C.W. Anderson, M.H. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, IEEE Trans. Neural Syst. Rehabilitation Eng. 11 (2003) 141–144.
[3] N. Kerkeni, F. Alexandre, M.H. Bedoui, L. Bougrain, M. Dogui, Automatic classification of sleep stages on a EEG signal by artificial neural networks, in: 5th WSEAS International Conference on SIGNAL, SPEECH and IMAGE PROCESSING-WSEAS SSIP'05.
[4] I.A. Aljazaery, A.A. Ali, H.M. Abdulridha, Classification of electroencephalograph (EEG) signals using quantum neural network, Signal Processing: An Int. J. (SPIJ) 4 (2011) 329.
[5] M. Nandish, M. Stafford, P. Kumar, F. Ahmed, Feature extraction and classification of EEG signal using neural network based techniques, Int. J. Eng. Innovative Technol. (IJEIT) 2 (2012) 1–5.
[6] B. Richhariya, M. Tanveer, EEG signal classification using universum support vector machine, Expert Syst. Appl. 106 (2018) 169–182.
[7] R. Palaniappan, Utilizing gamma band to improve mental task based brain-computer interface design, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2006) 299–303.
[8] L. Zhang, W. He, C. He, P. Wang, Improving mental task classification by adding high frequency band information, J. Med. Syst. 34 (2010) 51–60.
[9] R. Ameri, A. Pouyan, V. Abolghasemi, Eeg signal classification based on sparse representation in brain computer interface applications, in: Biomedical Engineering (ICBME), 2015 22nd Iranian Conference on, IEEE, pp. 21–24.
[10] A. Gupta, R.K. Agrawal, J.S. Kirar, J. Andreu-Perez, W.-P. Ding, C.-T. Lin, M. Prasad, On the utility of power spectral techniques with feature selection techniques for effective mental task classification in noninvasive BCI, IEEE Trans. Syst., Man, Cybern.: Syst. (2019).
[11] A. Gupta, R. Agrawal, B. Kaur, Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods, Soft. Comput. 19 (2015) 2799–2812.
[12] M. Tanveer, R.B. Pachori, N. Angami, Classification of seizure and seizure-free EEG signals using hjorth parameters, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, pp. 2180–2185.
[13] M. Tanveer, R.B. Pachori, N. Angami, Entropy based features in fawt framework for automated detection of epileptic seizure EEG signals, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, pp. 1946–1952.

[14] M. Hariharan, V. Vijean, R. Sindhu, P. Divakar, A. Saidatul, S. Yaacob, Classification of mental tasks using stockwell transform, Computers Electr. Eng. 40 (2014) 1741–1749.
[15] S. Siuly, Y. Li, Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (2012) 526–538.
[16] C. Xu, D. Tao, C. Xu, A survey on multi-view learning, arXiv preprint arXiv:1304.5634 (2013).
[17] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the eleventh annual conference on Computational learning theory, ACM, pp. 92–100.
[18] I. Muslea, S. Minton, C.A. Knoblock, Active + semi-supervised learning = robust multi-view learning, in: ICML, volume 2, Citeseer, pp. 435–442.
[19] V. Sindhwani, D.S. Rosenberg, An RKHS for multi-view learning and manifold co-regularization, in: Proceedings of the 25th international conference on Machine learning, ACM, pp. 976–983.
[20] K. Nigam, R. Ghani, Analyzing the effectiveness and applicability of co-training, in: Proceedings of the ninth international conference on Information and knowledge management, ACM, pp. 86–93.
[21] S. Sun, F. Jin, Robust co-training, Int. J. Pattern Recognit Artif Intell. 25 (2011) 1113–1126.
[22] I. Muslea, S. Minton, C.A. Knoblock, Active learning with multiple views, J. Artif. Intell. Res. 27 (2006) 203–233.
[23] H. Kadri, S. Ayache, C. Capponi, S. Koço, F.-X. Dupé, E. Morvant, The multi-task learning view of multimodal data, in: Asian Conference on Machine Learning, pp. 261–276.
[24] S. Sun, Multi-view Laplacian support vector machines, in: International Conference on Advanced Data Mining and Applications, Springer, pp. 209–222.
[25] S. Sun, J. Shawe-Taylor, Sparse semi-supervised learning using conjugate functions, J. Mach. Learn. Res. 11 (2010) 2423–2455.
[26] X. Xie, S. Sun, Multi-view twin support vector machines, Intell. Data Anal. 19 (2015) 701–712.
[27] X. Xie, S. Sun, Multi-view Laplacian twin support vector machines, Appl. Intell. 41 (2014) 1059–1068.
[28] P. Dhillon, D.P. Foster, L.H. Ungar, Multi-view learning of word embeddings via cca, in: Advances in neural information processing systems, pp. 199–207.
[29] T. Jaakkola, M. Meila, T. Jebara, Maximum entropy discrimination, in: Advances in neural information processing systems, pp. 470–476.
[30] H. Yanga, J. Heb, Notam^2: Nonparametric bayes multi-task multi-view learning, in: Proceedings of World Statistics Conference, pp. 2351–2356.
[31] X. Jin, F. Zhuang, S. Wang, Q. He, Z. Shi, Shared structure learning for multiple tasks with multiple views, in: Joint European conference on machine learning and knowledge discovery in databases, Springer, pp. 353–368.
[32] Z. Ding, Y. Fu, Low-rank common subspace for multi-view learning, in: Data Mining (ICDM), 2014 IEEE International Conference on, IEEE, pp. 110–119.
[33] M. Gönen, G.B. Gönen, F.S. Gürgen, Bayesian multiview dimensionality reduction for learning predictive subspaces., in: ECAI, pp. 387–392.
[34] P. Yang, W. Gao, Information-theoretic multi-view domain adaptation: a theoretical and empirical study, J. Artif. Intell. Res. 49 (2014) 501–525.
[35] B. Tan, E. Zhong, E.W. Xiang, Q. Yang, Multi-transfer: Transfer learning with multiple views and multiple sources, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM, pp. 243–251.
[36] Y. Makihara, A. Mansur, D. Muramatsu, Z. Uddin, Y. Yagi, Multi-view discriminant analysis with tensor representation and its application to cross-view gait recognition, in: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, volume 1, IEEE, pp. 1–8.
[37] S. Sun, X. Xie, M. Yang, Multiview uncorrelated discriminant analysis, IEEE transactions on cybernetics 46 (2016) 3272–3284.
[38] C. Zhang, H. Fu, S. Liu, G. Liu, X. Cao, Low-rank tensor constrained multiview subspace clustering, in: Proceedings of the IEEE international conference on computer vision, pp. 1582–1590.
[39] X. Cao, C. Zhang, H. Fu, S. Liu, H. Zhang, Diversity-induced multi-view subspace clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586–594.
[40] A. Anuragi, D.S. Sisodia, R.B. Pachori, Automated alcoholism detection using fourier-bessel series expansion based empirical wavelet transform, IEEE Sens. J. 20 (2020) 4914–4924.
[41] M. Sani, H. Norhazman, H. Omar, N. Zaini, S. Ghani, Support vector machine for classification of stress subjects using eeg signals, in: 2014 IEEE Conference on Systems, Process and Control (ICSPC 2014), IEEE, 2014, pp. 127–131.
[42] A. Bhattacharyya, R.B. Pachori, A. Upadhyay, U.R. Acharya, Tunable-q wavelet transform based multiscale entropy measure for automated classification of epileptic eeg signals, Appl. Sci. 7 (2017) 385.
[43] R. Sharma, R.B. Pachori, A. Upadhyay, Automatic sleep stages classification based on iterative filtering of electroencephalogram signals, Neural Comput. Appl. 28 (2017) 2959–2978.
[44] J.A. de la O Serna, M.R.A. Paternina, A. Zamora-Méndez, R.K. Tripathy, R.B. Pachori, Eeg-rhythm specific taylor-fourier filter bank implemented with o-splines for the detection of epilepsy using eeg signals, IEEE Sensors Journal (2020).
[45] H.U. Amin, W. Mumtaz, A.R. Subhani, M.N.M. Saad, A.S. Malik, Classification of eeg signals based on pattern recognition approach, Front. Comput. Neurosci. 11 (2017) 103.
[46] S. Sun, Q. Zhang, Multiple-view multiple-learner semi-supervised learning, Neural Processing Lett. 34 (2011) 229.
[47] L. Rokach, Pattern classification using ensemble methods, vol. 75, World Scientific, 2010.
[48] V.K. Singh, V. Kumar, A. Krishnamachari, Prediction of replication sites in saccharomyces cerevisiae genome using dna segment properties: multi-view ensemble learning (MEL) approach, Biosystems 163 (2018) 59–69.
[49] P. Lerman, Fitting segmented regression models by grid search, J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 29 (1980) 77–84.
[50] V. Kumar, S. Minz, Multi-view ensemble learning: an optimal feature set partitioning for high-dimensional data classification, Knowl. Inf. Syst. 49 (2016) 1–59.
[51] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693.
[52] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454 (1998) 903–995.
[53] R.B. Pachori, Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition, Res. Lett. Signal Process. 2008 (2008) 14.
[54] H. Berger, Über das elektrenkephalogramm des menschen, Archiv für psychiatrie und nervenkrankheiten 87 (1929) 527–570.
[55] P.F. Diez, V. Mut, E. Laciar, A. Torres, E. Avila, Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification, in: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, IEEE, pp. 2579–2582.
[56] A. Gupta, R. Agrawal, B. Kaur, A three phase approach for mental task classification using EEG, in: Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, pp. 898–904.
[57] N. Hazarika, J.Z. Chen, A.C. Tsoi, A. Sergejew, Classification of EEG signals using the wavelet transform, in: Digital Signal Processing Proceedings, 1997. DSP 97, 1997 13th International Conference on, volume 1, IEEE, pp. 89–92.
[58] J. Gilles, Empirical wavelet transform, IEEE Trans. Signal Processing 61 (2013) 3999–4010.
[59] I. Daubechies, Ten lectures on wavelets, Siam 61 (1992).
[60] J.C. Bezdek, R. Ehrlich, W. Full, Fcm: The fuzzy c-means clustering algorithm, Computers Geosci. 10 (1984) 191–203.
[61] L.A. Zadeh, A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, in: Systems Theory in the Social Sciences, Springer, 1976, pp. 202–282.
[62] H.T. Nguyen, A note on the extension principle for fuzzy sets, J. Math. Anal. Appl. 64 (1978) 369–380.
[63] P.K. Tiwari, A.K. Srivastava, Zadeh extension principle: a note, Ann Fuzzy Math Inform 9 (2014) 37–41.
[64] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inform. Theory 22 (1976) 75–81.
[65] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Mining Knowl. Discovery 2 (1998) 121–167.
[66] G.N. Garcia, T. Ebrahimi, J.-M. Vesin, Support vector EEG classification in the fourier and time-frequency correlation domains, in: Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on, IEEE, pp. 591–594.
[67] B. Blankertz, G. Curio, K.-R. Müller, Classifying single trial EEG: Towards brain computer interfacing, in: Advances in neural information processing systems, pp. 157–164.
[68] R. Kumari, J. Jose, Seizure detection in EEG using biorthogonal wavelet and fuzzy KNN classifier, Elixir Hum. Physiol 41 (2011) 5766–5770.
[69] S.B. Imandoust, M. Bolandraftar, Application of k-nearest neighbor (KNN) approach for predicting economic events: Theoretical background, Int. J. Eng. Res. Appl. 3 (2013) 605–610.
[70] B.V. Dasarathy, Nearest neighbor (NN) norms:NN pattern classification techniques(1991).
[71] T. Fawcett, An introduction to ROC analysis, Pattern Recog. Lett. 27 (2006) 861–874.
[72] A.T. Azar, S.A. El-Said, Performance analysis of support vector machines classifiers in breast cancer mammography recognition, Neural Comput. Appl. 24 (2014) 1163–1177.
[73] Z.A. Keirn, J.I. Aunon, A new mode of communication between man and his surroundings, IEEE Trans. Biomed. Eng. 37 (1990) 1209–1214.
[74] N.-J. Huan, R. Palaniappan, Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals, in: Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on, IEEE, pp. 633–636.
[75] C.W. Anderson, E.A. Stolz, S. Shamsunder, Discriminating mental tasks using EEG represented by AR models, in: Engineering in Medicine and Biology Society, 1995, IEEE 17th Annual Conference, volume 2, IEEE, pp. 875–876.
[76] S. Rezaei, K. Tavakolian, K. Naziripour, Comparison of five different classifiers for classification of mental tasks, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, pp. 6007–6010.
[77] L. Zhiwei, S. Minfen, Classification of mental task EEG signals using wavelet packet entropy and SVM, in: 2007 8th International Conference on Electronic Measurement and Instruments, IEEE, pp. 3–906.

Dr. Akshansh Gupta is currently working as a scientist in Central Electronics Engineering Research Institute Pilani, Rajasthan, 333031, India. He is also working as a principal investigator in a DST funded scheme of Cognitive Science Research Initiative (CSRI) from the Department of Science and Technology (DST), Ministry of Science and Technology, Government of India, with the project entitled "Identification of EEG signature for different mental states in School of Computational Integrative and Science, Jawaharlal Nehru University, New Delhi. He received his master and Ph.D. degree from School of computer and systems sciences, JNU, in the year 2010 and 2015 respectively. His research interests include Pattern Recognition, Machine Learning, Data Mining Signal Processing, Brain Computer Interface, Cognitive Science, IoT. He is also working a as CO-PI on a consultancy project named as "Development of Machine Learning Algorithms for Automated Classification Based on Advanced Signal Decomposition of EEG Signals" ICPS Program, DST Govt. Of India.

R.U. Khan is working as Senior Research Fellow (SRF) in Discipline of Mathematics, Indian Institute of Technology, Indore, India. He received his B.Tech. and M.Tech. degree from Aligarh Muslim University, Aligarh, India, in 2015 and 2017 respectively. His research interest includes bio medical signal processing, machine learning, brain computer interface, neurodegenerative diseases, computer-aided medical diagnosis, and Alzheimer's disease (AD).

V. Singh received his B.Tech. in Biotechnology Engineering from Kurukshetra University in 2007, and his MTech. and Ph.D. in Computational Biology from Jawaharlal Nehru University, New Delhi, in 2010 and 2017 respectively, where he applied machine learning and data analysis techniques to get insights into the biological datasets.

M. Tanveer is Associate Professor and Ramanujan Fellow at the Discipline of Mathematics of the Indian Institute of Technology, Indore. Prior to that, he spent one year as a Postdoctoral Research Fellow at the Rolls-Royce@NTU Corporate Lab of the Nanyang Technological University, Singapore. His research interests include machine learning, deep learning, optimization, biomedical engineering and applications to Alzheimer's disease and dementias. He has published over 40 referred journal papers of international repute. He is the recipient of the 2016 DST-Ramanujan Fellowship in Mathematical Sciences and 2017 SERB-Early Career Research Award in Engineering Sciences which are the prestigious awards of INDIA at early career level. He is a Senior Member of IEEE, editorial review board member of Applied Intelligence, Springer, Guest Editor of ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Multimedia Tools and Applications, Springer, Applied Soft Computing, Elsevier and Associate Editor for IEEE SMC 2019. He has also co-edited one book in Springer on machine intelligence and signal analysis. He has also been lead organizer/general chair and invited/plenary/keynote speaker in many international conferences and Symposiums. He was the Co-Chair of Special Session Proposal in 2018 IEEE SSCI 2018. Tanveer is currently the Principal Investigator of 06 research projects, and Co-Principal Investigator of one international project (MHRD SPARC Scheme) funded by Government of India including Department of Science and Technology (DST), Science & Engineering Research Board (SERB), Council of Scientific & Industrial Research (CSIR), Ministry of Human Resource Development (MHRD).

D. Kumar is working as Assistant Professor in Department of Applied Mathematics, Delhi Technological University, Delhi, India. He is pursuing Ph.D degree from Jawaharlal Nehru University, New Delhi. He received his B.Sc. degree from Banaras Hindu University, Varanasi, year 2011. He had completed M.Sc. in computer science from the same university in the year 2013. He has done M. Tech. in Computer Science and Technology from Jawaharlal Nehru University in year 2015. His research interest includes medical image segmentation, pattern recognition and image classification.

A. Chakraborti is a Professor at the School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, since March, 2014. He had worked as an Associate Professor at the Chair of Quantitative Finance, Ecole Centrale Paris, France, during 2009–14. He had obtained a Ph.D. in Physics from Saha Institute of Nuclear Physics, India and then the Habilitation in Physics from Universite Pierre et Marie Curie (Paris VI), France. He has the experience of working as a scientist in many reputed universities and educational institutions in India, USA, Europe and Japan. He was awarded the prestigious Young Scientist Medal of the Indian National Science Academy in 2009. He has published several books and research articles from internationally renowned publishers. His main research interests lie in the areas of Econophysics, Sociophysics, Data Science, Statistical Physics, Quantum Physics and Nanomaterial Science.

Ram Bilas Pachori received the B.E. degree with honours in Electronics and Communication Engineering from Rajiv Gandhi Technological University, Bhopal, India in 2001, the M.Tech. and Ph.D. degrees in Electrical Engineering from Indian Institute of Technology (IIT) Kanpur, Kanpur, India in 2003 and 2008, respectively. He worked as a Postdoctoral Fellow at Charles Delaunay Institute, University of Technology of Troyes, Troyes, France during 2007–2008. He served as an Assistant Professor at Communication Research Center, International Institute of Information Technology, Hyderabad, India during 2008–2009. He served as an Assistant Professor at Discipline of Electrical Engineering, IIT Indore, Indore, India during 2009–2013. He worked as an Associate Professor at Discipline of Electrical Engineering, IIT Indore, Indore, India during 2013–2017 where presently he has been working as a Professor since 2017. He is also an Associated Faculty with Discipline of Biosciences & Biomedical Engineering at IIT Indore. He was a Visiting Professor at the School of Medicine, Faculty of Health and Medical Sciences, Taylors University, Subang Jaya, Malaysia during 2018–2019. He worked as a Visiting Scholar at Intelligent Systems Research Center, Ulster University, Northern Ireland, UK during December 2014. He is an Associate Editor of Electronics Letters, Biomedical Signal Processing and Control journal and an Editor of IETE Technical Review journal. He is a senior member of IEEE and a Fellow of IETE. He has more than 190 publications which include journal papers, conference papers, books, and book chapters. His publications have around 6000 citations, h index of 41, and i10 index of 98 (Google Scholar, May 2020). He has served on review boards for around 100 scientific journals and served for scientific committees of various national and international conferences. His research interests are in the areas of biomedical signal processing, non-stationary signal processing, speech signal processing, signal processing for communications, computer-aided medical diagnosis, and signal processing for mechanical systems.