IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 13, 2019, accepted February 22, 2019, date of publication March 12, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904400

Internal Emotion Classification Using EEG Signal With Sparse Discriminative Ensemble

HABIB ULLAH[1], MUHAMMAD UZAIR[2], ARIF MAHMOOD[3], MOHIB ULLAH[4], SULTAN DAUD KHAN[1], AND FAOUZI ALAYA CHEIKH[4]

[1]College of Computer Science and Engineering, University of Hail, Ha'il 55476, Saudi Arabia
[2]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan
[3]Department of Computer Science, Information Technology University, Lahore 54000, Wah Cantonment 47040, Pakistan
[4]Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding author: Mohib Ullah (mohib.ullah@ntnu.no)

This work was supported by the Norwegian University of Science and Technology through the open access journal publication fund.

ABSTRACT Among various physiological signal acquisition methods for the study of the human brain, EEG (Electroencephalography) is more effective. EEG provides a convenient, non-intrusive, and accurate way of capturing brain signals in multiple channels at fine temporal resolution. We propose an ensemble learning algorithm for automatically computing the most discriminative subset of EEG channels for internal emotion recognition. Our method describes an EEG channel using kernel-based representations computed from the training EEG recordings. For ensemble learning, we formulate a graph embedding linear discriminant objective function using the kernel representations. The objective function is efficiently solved via sparse non-negative principal component analysis and the final classifier is learned using the sparse projection coefficients. Our algorithm is useful in reducing the amount of data while improving computational efficiency and classification accuracy at the same time. The experiments on publicly available EEG dataset demonstrate the superiority of the proposed algorithm over the compared methods.

INDEX TERMS Multiple channel EEG, emotion recognition, linear discriminant analysis, sparse PCA.

I. INTRODUCTION

Emotions are an important aspect of human communication and decision making [1]. For computer based analysis of human emotions, different physiological signal measurement methods such as Electromyography (EMG) [2], Electrocardiography (ECG) [3], respiration rate, galvanic skin response and Electroencephalography (EEG) [4] have been used. Among these, EEG is more effective since it provides convenient, non-intrusive and more accurate way of capturing brain signals. Multiple channel EEG signals encapsulate important emotional clues of human brain dynamics at finer temporal resolution. Moreover, methods from the rich signal processing literature can be readily applied to EEG brain signals for the development of new approaches to effective computing [5].

Emotion recognition from EEG signals has recently attracted significant research attention [4], [7]–[10]. The reasons include its wide scope applications and the latest developments in portable and low cost EEG devices (Fig.1).

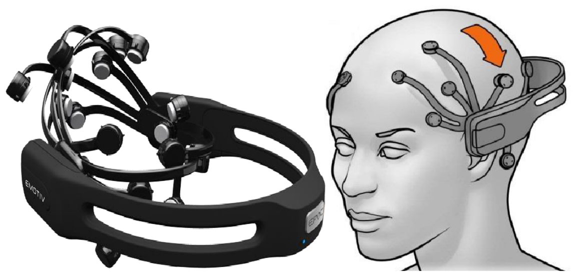


FIGURE 1. **Low cost multiple channel portable wireless EEG device from Emotiv [6].**

Internal emotion recognition systems have applications in many diverse areas including human-computer interaction, emotion understanding, brain-computer interface, and healthcare [11]–[13]. For example, the EEG signals can be used in real-time to detect emotions and the mental states including concentration levels. This information can be used as a feedback to activate different actions in technologically advanced applications, e.g. to change a scene in a virtual reality environment or refine lecture delivery in E-learning system. The

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

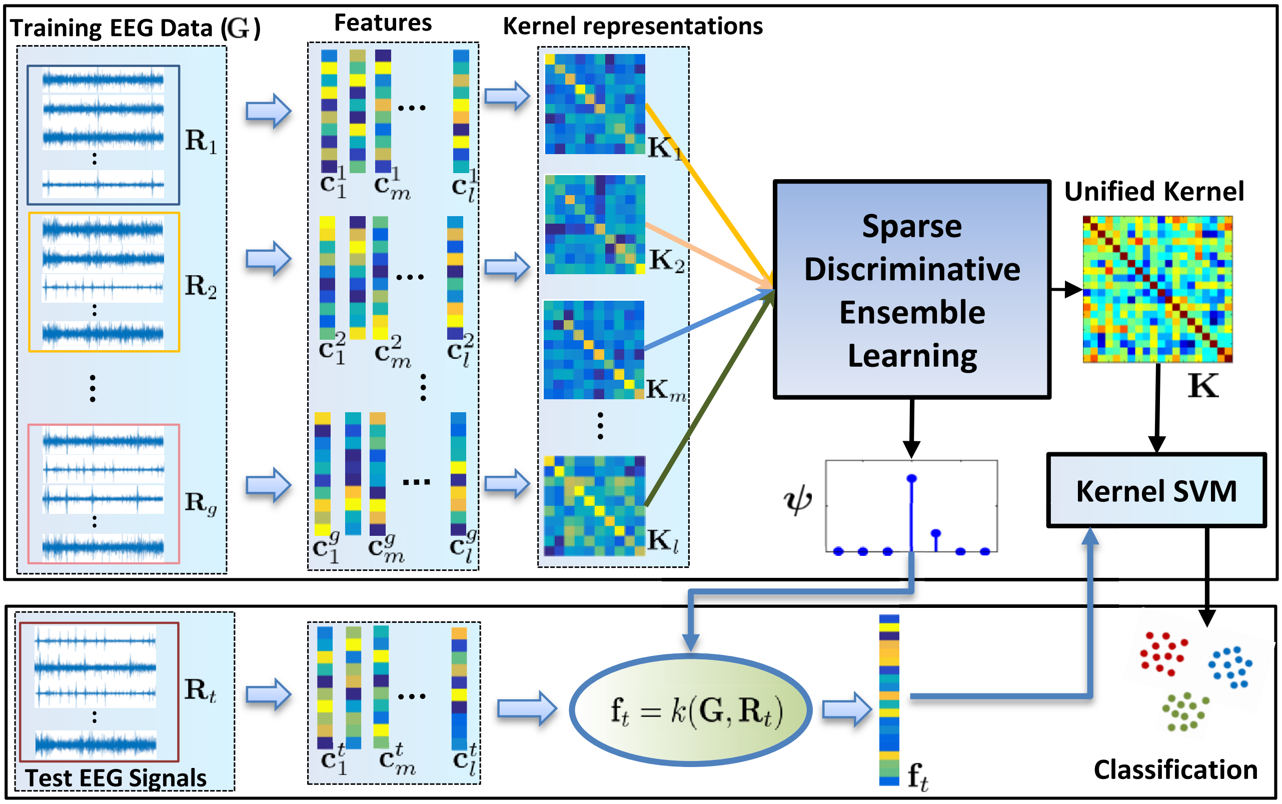


FIGURE 2. Illustration of the proposed sparse channel ensemble learning framework. Each channel i of the input EEG recording $\mathbf{R}_j$ is represented as statistical and frequency domain features denoted by c_i^j. During training, SDEL algorithm computes a bag of basic kernels from different channels and then automatically learns the best ensemble kernel from their sparse superposition. In the test stage, only the selected most discriminative kernel representations are used to compute features for classification.

detection of emotional states is also important for clinical applications especially for patients with disabilities who cannot directly communicate. Moreover, the fundamental characteristics of emotions can also help psychiatrists to study psychological disorders such as anxiety and depression [14].

EEG based automatic emotion recognition [7], [15] is a challenging pattern recognition problem because of the vague boundaries and differences in individual signatures of emotions. EEG data represents brain waves which are acquired using electrodes placed on scalp according to a specific standard such as the international 10-20 system [16]. The acquired EEG data is captured in multiple channels corresponding to multiple electrodes. The rhythmic activity in EEG signals is described in different frequency bands such as delta (<4Hz), theta (4-7Hz), alpha (8-13Hz), beta (14-30 Hz) and gamma (>30Hz). There are two important steps in the design of an emotion recognition system: the extraction of effective features characterized by discriminant information represented by small compact values, and the design of accurate learning algorithms for the classification of these features. Different feature extraction and classification exist in the literature for multiple channel EEG based emotion recognition.

In multiple channel EEG recordings all EEG channels may not be equally discriminative for the task of emotion recognition. Therefore, methods considering all channels equal in classifier learning suffer from redundancy in the data. To reduce the amount of data, state-of-the-art methods consider channel selection for improving emotion recognition accuracy [17]. Inspired by the channel selection and sparse coding [18]–[22] literature, in this paper, we propose a Sparse Discriminative Ensemble Learning (SDEL) algorithm for multiple channel EEG based emotion recognition (Fig. 2). The SDEL learns a subset of the most discriminative channels automatically from the available training EEG recordings in a supervised setting. Our method is generic and is applicable as a pre-processing step of any multiple channels EEG based recognition system.

The main contributions of this paper are:

1) Given multiple channel EEG recordings which represent different classes of emotions, we efficiently describe individual channels using Radial Basis Functions (RBF) kernel representations.
2) Using our kernel representations of the training data, we formulate a new ensemble learning method via a supervised graph embedding linear discriminant analysis objective function.
3) We propose an efficient solution to the objective function using a sparse non-negative principal component analysis algorithm to find the most dominant projection

which minimizes the within-class scatter and maximizes the between-class scatter. The coefficients of the projection vector are then used as the weights of the individual channels in the final classifier learning.

4) We show that our proposed SDEL can be used as a pre-processing method to improve the accuracy of different EEG based emotion recognition algorithms.

The proposed SDEL is useful in significantly reducing the amount of data for learning the final classifier while improving the efficiency and accuracy at the same time. Moreover, our algorithm does not require expensive validation data to learn the parameters. After learning an optimal composite of channels and their relative importance, we input these to the emotion classification system. We evaluate the effectiveness of our proposed method on benchmark EEG dataset. The proposed method improves the accuracy of emotion recognition using multiple channel EEG recordings over the compared algorithms.

II. RELATED WORK

We categorize EEG based emotion recognition literature into biologically-inspired methods, wavelet based methods and deep learning based methods. We also give detail review of the feature and channel selection methods.

A. BIOLOGICALLY INSPIRED METHODS

Zhuang *et al.* [23] and Khosrowabadi *et al.* [24] introduced biologically-inspired algorithms for emotion recognition. Zhuang *et al.* [23] investigated empirical mode decomposition to break up the EEG signals into intrinsic mode functions (IMF). The IMF is multidimensional information that is considered as features for emotion recognition. Khosrowabadi *et al.* [24] employed a neural network (feedforward) with a shift memory register and spectral filtering for human emotion recognition from EEG signals. Sreeshakthy and Preethi [25] selected different features using particle swarm optimization. Their classification method was based on the radial basis function networks.

B. WAVELET BASED METHODS

In the wavelet based methods, Fernández-Varela *et al.* [4] decomposed EEG signals into corresponding frequency bands and extracted several features. SVM is then used to detect emotion states. Jenke *et al.* [15] considered a set of multiple features of EEG signals. Murugappan [26] used three wavelet functions to extract statistical features and recognized emotions considering KNN classifier. Jalilifard *et al.* [27] pre-processed the EEG signals by stationary wavelet transform and then computed the distribution of power in time-frequency space. Then 46 features were extracted and classified using SVM. Murugappan *et al.* [28] used time-frequency analysis of wavelet transform and surface Laplacian filtering with linear classifiers for emotion recognition.

C. DEEP-LEARNING BASED METHODS

In the deep learning based methods, Mehmood *et al.* [29] introduced deep learning ensemble method to select optimal combination of features for emotion recognition. Mohammad *et al.* [30] proposed long-short-term-memory recurrent-neural networks and continuous conditional random fields to detect and classify emotions. Gao *et al.* [31] proposed a deep learning technique using restricted Boltzmann machines. They simultaneously learned the features and classifier from raw EEG signals. Weninger *et al.* [32] performed multi-variate regression by deep recurrent-neural networks to model longer-range context and capture the time-varying emotional profile of musical pieces. Tripathi *et al.* [33] proposed deep and convolutional neural networks for emotion recognition from multi-channel EEG signals. They extracted simple statistical features from each channel and then trained deep neural network models for emotion classification in two and three states. Similarly, Song *et al.* [34] used dynamical graph convolutional neural networks for feature learning from EEG signals.

D. FEATURE AND CHANNEL SELECTION METHODS

We categorize channel combination and selection techniques into filtering, embedded and hybrid methods. The filtering methods exploit an evaluation criterion based on some distance and/or information metric to assess a subset of channels generated using a search algorithm. In the embedded methods, a few discriminative channels are selected based on selection criterion in conjunction with classifier learning objective function. The hybrid methods combine different techniques to take their collective advantages and avoid pre-specification of stopping criteria.

1) FILTERING METHODS

In the filtering methods, Al-Ani and Mesbah [35] used a simple search technique using different classifiers such as SVM and Extreme Learning Machines to identify the best performing channels and feature in a two stage processes. Ackermann *et al.* [36] used a filtering technique to select and combine discriminative channels and classify a set of emotions. Masood *et al.* [37] introduced a variant of common spatial pattern algorithm to select least number of EEG channels. They identified the spatial filter weights using complete set of channels and selected channels based on the maximal filter weights.

2) EMBEDDED METHODS

In the embedded methods, Zheng [38] extended the conventional canonical correlation analysis algorithm to model the linear correlation between class labels and the corresponding EEG feature vectors. They used group feature selection to simultaneously cope with both automatic channel selection and emotion recognition. Ansari-Asl *et al.* [39] proposed synchronization likelihood (SL) in multivariate data sets for channel selection.

3) HYBRID METHODS

In the hybrid methods, Mehmood *et al.* [29] selected and combined EEG channels using a balanced one-way ANOVA (analysis of variance) which were then classified using support vector machine, k-nearest neighbor and linear discriminant analysis. Zhang *et al.* [40] used the Relief algorithm combined with SVM to select features for emotion recognition. Polat and Güneş [41] used fast Fourier transform based features and decision trees to select and combine EEG channels.

The proposed Sparse Discriminative Ensemble Learning (SDEL) algorithm falls in the embedded category and automatically learns the most discriminative subset of channels in a supervised framework without the use of validation set. As SDEL uses a linear discriminant analysis based criterion, it is more appropriate for classification oriented ensemble learning. Moreover, SDEL is more efficient compared to the methods which involve more expensive sequential backward/forward evaluation/selection strategies.

III. LEARNING A SPARSE DISCRIMINATIVE COMBINATION OF EEG CHANNELS

We present a Sparse Discriminative Ensemble Learning (SDEL) algorithm for multiple channel EEG based emotion recognition. Considering supervised setting, our SDEL algorithm learns a subset of the most discriminative channels automatically from the available training EEG recordings. In this section, we present kernel based representation of channels, objective function, optimization, and kernel support vector machine based classification.

A. KERNEL BASED REPRESENTATION OF CHANNELS

Let $\mathbf{R}_j = [\mathbf{c}_1^j, \mathbf{c}_2^j, \ldots, \mathbf{c}_l^j] \in \mathbb{R}^{d\times l}$, is the data matrix of the j-th EEG recording containing l channels as its columns, where $\mathbf{c}_m^j \in \mathbb{R}^d$ denotes the m-th channel described with a d dimensional feature representation. In this work, l and d are same across all EEG recordings. Let $\mathbf{G} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_g\}$ be the training EEG signal recordings belonging to c different classes with labels $\mathbf{Y} = \{y_1, y_2, \ldots, y_g\}$.

A straightforward approach is to use each channel individually for learning l different classification models and then combine their scores at the decision level. However, learning l classifiers is time and resource inefficient. Moreover, finding the best combination parameters requires expensive validation data which is also not available most of the times. In contrast, we propose a more effective way to encode the discriminative ability of the same channel of different EEG recording via kernel representations. Our method works on the available training data and does not use any validation set for parameter tuning. Specifically, using the Gaussian radial basis function, we define l kernel matrices $\{\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_l\}$ such that a kernel matrix $\mathbf{K}_m \in \mathbb{R}^{g\times g}$ is defined by:

$$\mathbf{K}_m(i,j) = k_m(\mathbf{R}_i, \mathbf{R}_j) = \exp\left(\frac{-d^2(\mathbf{c}_m^i, \mathbf{c}_m^j)}{\sigma^2}\right) \tag{1}$$

where $d(\mathbf{c}_m^i, \mathbf{c}_m^j)$ is a distance measure (usually Euclidean distance) between channels $\mathbf{c}_m^i$ and $\mathbf{c}_m^j$ and σ is the kernel parameter.

We next formulate the problem of discriminative channel selection as a sparse multiple kernel learning problem over the kernel matrices. Our kernel based approach can efficiently handle the high dimensionality of EEG recordings. Moreover, an optimal subset of kernels (and thus channels) will be learned automatically according to their discriminative ability.

B. OBJECTIVE FUNCTION

We approach the kernel learning for channel selection problem by formulating it as graph embedding linear discriminant analysis objective function using the training kernel representations. Furthermore, to select only a few channels having the highest discriminative power, we embed sparsity in the formulation. We define g tensor matrices $\{\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_g\}$, where $\mathbf{\Lambda}_i = [\mathbf{K}_1(i), \mathbf{K}_2(i), \ldots, \mathbf{K}_l(i)] \in \mathbb{R}^{g\times l}$ and $\mathbf{K}(i)$ is the i-th column of $\mathbf{K}$. Next, we want to perform discriminant analysis on these tensors. In doing so, the sample coefficients of each tensor will be learned in a discriminative manner and thus can be used to linearly combine channels for improved recognition accuracy. To formulate our problem using graph embedding discriminant analysis, we can represent the itra-class scatter ($\boldsymbol{\zeta}_w$) and the inter-class scatter ($\boldsymbol{\zeta}_b$) using:

$$\boldsymbol{\zeta}_w = \sum_{i,j=1}^{g} z_{ij}(\mathbf{\Lambda}_i - \mathbf{\Lambda}_j)^\top(\mathbf{\Lambda}_i - \mathbf{\Lambda}_j), \tag{2}$$

$$\boldsymbol{\zeta}_b = \sum_{i,j=1}^{g} \acute{z}_{ij}(\mathbf{\Lambda}_i - \mathbf{\Lambda}_j)^\top(\mathbf{\Lambda}_i - \mathbf{\Lambda}_j). \tag{3}$$

with $z_{ij} = \begin{cases} 1/n_k & \text{if } (\mathbf{\Lambda}_i, \mathbf{\Lambda}_j) \in c_k, \\ 0 & \text{otherwise,} \end{cases}$, $\acute{z}_{ij} = 1/g$ and n_k represents the total training EEG recordings in class c_k having label y_k. The objective function is then defined as:

$$\boldsymbol{\psi}^* = \underset{\|\boldsymbol{\psi}\|_2=1}{\arg\max} \quad \boldsymbol{\psi}^\top(\boldsymbol{\zeta}_b - \boldsymbol{\zeta}_w)\boldsymbol{\psi} \tag{4}$$

In the above objective function, maximization of the scatter difference is introduced. Therefore, the optimal solution $\boldsymbol{\psi}^*$ will be able to minimize $\boldsymbol{\zeta}_w$ (intra-class scatter) and maximize $\boldsymbol{\zeta}_b$ (inter-class scatter). We further impose sparsity in the objective function for sparse channel selection and are interested in only the most dominant projection direction. Thus, our objective function becomes:

$$\boldsymbol{\psi}^* = \underset{\boldsymbol{\psi}\in\mathbb{S}_k^l}{\arg\max} \quad \boldsymbol{\psi}^\top(\boldsymbol{\zeta}_b - \boldsymbol{\zeta}_w)\boldsymbol{\psi} \tag{5}$$

where $\mathbb{S}_s^l = \{\boldsymbol{\psi} \in \mathbb{R}^l : \|\boldsymbol{\psi}\|_2 = 1, \|\boldsymbol{\psi}\|_0 \leq s, \boldsymbol{\psi} \geq 0\}$, for the desired sparsity $s \in [l]$.

C. OPTIMIZATION

We solve the objective function in Eq. 5 by the Non-negative Sparse Principal Component Analysis (NNSPCA) algorithm of Asteris *et al.* [42]. The detail of the algorithm for finding the most dominant non-negative sparse principal component (PC) is given below.

Let $\mathbf{M} = \boldsymbol{\zeta}_b - \boldsymbol{\zeta}_w$ be our PSD matrix, s be the desired sparsity level, and $r \in [l]$ be the accuracy parameter. The NNSPCA algorithm computes a non-negative, s-sparse, unit norm vector $\boldsymbol{\psi}_d$ approximating the nonnegative, s-sparse PC of $\mathbf{M}$ using the following algorithm.

First, the rank-r approximation of $\mathbf{M}$ denoted by $\mathbf{M}_r$ is computed. $\mathbf{M}_r$ can be computed as the best rank-r approximation of $\mathbf{M}$ by zeroing out the $l - r$ trailing eigenvalues of $\mathbf{M}$, that is, $\mathbf{M}_r = \sum_i^r \alpha_i \mathbf{v}_i \mathbf{v}_i^\top$; where α_i is the i^{th} largest eigenvalue of $\mathbf{M}$ with the corresponding eigenvector $\mathbf{v}_i$.

Secondly, a set of $O(l^r)$ candidate supports denoted by S_r is computed. However, enumerating the $\binom{l}{s}$ possible supports for s-sparse vectors in $\mathbb{R}^l$ is computationally very expensive. Therefore, a spannogram algorithm [42] is used to efficiently compute a collection S_r of support sets with cardinality $|S_r| \leq 2^r\binom{l+1}{r}$. This has been shown to provably contain the support of the non-negative, s-sparse principal component of $\mathbf{M}_r$ [42]. Next, a set of candidate solutions denoted by Ψ_r is computed. Specifically, for each candidate support set $I \in S_r$, a candidate solution $\boldsymbol{\psi}$ supported only in I is computed by solving:

$$\underset{\|\boldsymbol{\psi}\|_2=1, \boldsymbol{\psi}\geq 0, supp(\boldsymbol{\psi})\subseteq I}{\arg\max} \quad \boldsymbol{\psi}^\top \mathbf{M_r} \boldsymbol{\psi} \tag{6}$$

We adopt the method of Asteris *et al.* [42] for solving the constrained quadratic maximization problem in Eq. 6. The best candidate solution in Ψ_r is selected which is the candidate that maximizes the quadratic objective function in (6). This solution is our required coefficient vector $\boldsymbol{\psi} \in \mathbb{R}^l$ which encodes information about the selected subset of channels and their relative strengths. In other words, the coefficients in $\boldsymbol{\psi}$ corresponds to the importance of the individual channels. The selected subset of channels together with $\boldsymbol{\psi}$ can now be used as input to an EEG emotion recognition algorithm where they can be exploited in combination to achieve improved accuracy at lower cost than using full set of channels.

D. KERNEL SUPPORT VECTOR MACHINE (SVM) BASED CLASSIFICATION

Each channel kernel computed via Eq. (1) is symmetric and can also be made semi-positive definite by adding to its diagonal a small positive constant. These valid kernels can then be used with a kernel based learning algorithm for classification. Furthermore, according to the Reproducing Kernel Hilbert Space (RKHS) theory [43], a linear superposition of valid kernels is a new valid kernel. Therefore, we can compute a new unified kernel using the kernel function:

$$k(\mathbf{R}_i, \mathbf{R}_j) = \sum_{m=1}^{l} \boldsymbol{\psi}(m) k_m(\mathbf{R}_i, \mathbf{R}_j), \tag{7}$$

where $\boldsymbol{\psi}$ is our learned sparse coefficient vector.

Support Vector Machine (SVM) is classical supervised learning method for two class classification problems in the Euclidean spance [44]. We adopt its kernel version by embedding the channel representation into the RKHS space [43].

Recall our training EEG signal recordings $\mathbf{G} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_g\}$ belonging to c different classes with labels $\mathbf{Y} = \{y_1, y_2, \ldots, y_g\}$ where $\mathbf{R}_j \in \mathbb{R}^{d\times l}$ is the data matix of an EEG recording containing channel features in its columns. For two class problems $y_i \in \{-1, +1\}$. SVM finds a maximum-margin hyperplane that optimally separates examples having $y_i = -1$ from the examples having $y_i = +1$. The parameters $\mathbf{w}, b$ of the optimal hyperplane are computed by solving the following soft-margin SVM optimization problem [44]:

$$\begin{aligned} \underset{\mathbf{w},b,\eta}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{g} \eta_i \\ \text{subject to} \quad & y_i\left(\mathbf{w}^T \mathbf{f}_i + b\right) \geq 1 - \eta_i, \quad i = 1, \ldots, g \\ & \eta_i \geq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{8}$$

where $\mathbf{w}$, b denote the parameters of the hyperplane, paramter η is used to accommodate the non separable cases and C is the penalty parameter [44]. $\mathbf{f}_i \in \mathbb{R}^g$ is the feature vector of $\mathbf{R}_i$ computed in the high dimensional RKHS space by using our unified kernel function of Eq. (7) as

$$\mathbf{f}_i = k(G, \mathbf{R}_i) \tag{9}$$

The following dual of the above convex problem is obtained using the Lagrangian duality [44]:

$$\begin{aligned} \underset{\beta}{\text{minimize}} \quad & \frac{1}{2}\sum_{i,j} \beta_i\beta_j y_i y_j k(\mathbf{R}_i, \mathbf{R}_j) - \sum_i \beta_i \\ \text{subject to} \quad & 0 \leq \beta_i \leq C, \quad i = 1, \ldots, m \\ & \sum \beta_i y_i = 0 \end{aligned} \tag{10}$$

This is a quadratic programming problem and is solved for optimal $\boldsymbol{\beta}^*$. The parameter of the hyperplane is then given in terms of $\boldsymbol{\beta}^*$, as $\mathbf{w}_{opt} = \sum_i \beta_i y_i \mathbf{f}_i$. After finding $\mathbf{w}_{opt}$, the intercept term b is calculated from the primal form.

Given a test EEG recording $\mathbf{R}_t$, we first compute its kernel domain feature representation $\mathbf{f}_t$ using Eq. 9, then using the learned SVM hyperplane, the label is predicted as:

$$y_t = \text{sign}\left(\frac{\mathbf{w}_{opt}^T \mathbf{f}_t + b}{\|\mathbf{w}^*\|}\right) \tag{11}$$

where function sign returns the sign of its argument.

IV. FEATURE EXTRACTION FOR EEG SIGNAL REPRESENTATION

We extract multiple statistical and frequency domain features from each channel of the EEG recording to represent it compactly as d dimensional feature vector. Given a EEG signal $\mathbf{c}$ having N samples, we compute different features as described next.

A. STATISTICAL FEATURES OF RAW SIGNALS

For statistical feature extraction, we first divide a given N sample EEG channel, into t segments. For each segment we compute its statistical mean, median, maximum, minimum, standard deviation, variance, range, skewness, kurtosis, Petrosian Fractal Dimension [45], [46], Fisher Information Ratio [45], [47] and entropy values. We also compute these feature values using all N samples of a channel. Finally, we concatenate these simple features to obtain the final statistical feature representation of a channel.

B. FREQUENCY DOMAIN FEATURES

To make the feature representation robust, we further extract multiple features by first transforming the signal to frequency domain and then extracting the following basic features.

1) STFT BASED FEATURES

We compute Power Spectral Density (PSD) and Differential Entropy (DE) [48] features using Short Time Fourier Transform (STFT) with a 1-s-long window and no overlapping in four frequency bands theta (4-7Hz), alpha (8-13Hz), beta (14-30Hz) and gamma (31-45Hz). Although, these features can be computed for the individual segments but we only consider computing them for the whole channel.

2) DISCRETE COSINE TRANSFORM FEATURES

The Discrete Cosine Transform (DCT) [49] expresses a discrete signal as a linear combination of mutually uncorrelated cosine basis functions. The advantage of DCT is that only a few transform coefficients optimally represent the signal information in a compact manner. Moreover, the DCT coefficients are real numbers and thus are efficient to process for feature representation. From the input signal, DCT computes the energy spectrum by:

$$\mathbf{F}(u) = \sum_{n=0}^{N-1} \mathbf{c}(n)\cos\left[\frac{\pi}{N}(n+\frac{1}{2})u\right] \quad u = 0, \ldots, N-1 \tag{12}$$

After computing the DCT, we select only a few low frequency DCT coefficient for feature representation. We compute DCT features for both the individual segments as well as for the whole channel. The final representation is obtained by concatenating the selected DCT coefficients of the segments with those of the whole channel.

3) SPECTROGRAM BASED FEATURES

We first divide the given input channel into multiple segments of equal length. Next, we compute STFT of each segment using the equation:

$$\text{STFT} = \mathbf{c}(n)(m, w) = \sum_{N=-\inf}^{\inf} \mathbf{c}(n)w(n-m)\exp^{-jwn} \tag{13}$$

where $\mathbf{c}(n)$ denotes the data segment and $w(n)$ is the window function. To compute the feature vector, we find the local maxima in each segment. The harmonic relationships of the detected maxima are then grouped together to form a feature vector. In order to determine the harmonic relations of the spectra peaks, each detected peak is assumed to be the fundamental in turns. After finding the harmonic of the current fundamental from the remaining peaks, the amplitudes of the strong harmonic set can be seen as the harmonic feature vector set. In order to minimize the influence of propagation distance, the feature vector is normalized by the magnitude of the highest harmonics. Finally the feature vectors from each segment are statistically averaged to form a spectrogram based feature vector.

V. EXPERIMENTS AND RESULTS

We perform extensive experiments on publicly available dataset using extensively in emotion recognition research. We also compare the performance of the proposed SDEL with other methods available in the literature. Furthermore, we experimentally evaluate the proposed SDEL by applying it as a pre-processing stage of different Emotion recognition algorithms and comparing the accuracy gain.

A. DATASET DESCRIPTION

DEAP [8] is a benchmark dataset widely used in the EEG based emotion recognition research. DEAP dataset contains EEG and peripheral physiological signals of 32 participants while they watched 40 examples of one-minute duration music videos. 32 active electrodes (channels) were used to record the EEG signals. The electrodes were placed on the head scalp by following the 10-20 international standard. Another 8 channels comprising the peripheral physiological signals including the skin temperature, blood volume pressure, galvanic skin response, respiration rate, electromyogram and electrooculogram (horizontal and vertical) were also recorded.

In our experiments, we use the pre-processed version of DEAP EEG data (Table 1). The pre-processing consists of downsampling to 128Hz, EOG removal and bandpass filtering (4.0-45.0Hz). This version is widely used to test classification and regression algorithms in the literature. The data is provided as $40 \times 40 \times 8064$ dimensional Matlab and Python arrays representing (trial$\times$channel$\times$data) for each of the 32 subjects. Similarly the labels are also provided as 40×4 arrays representing trial$\times$label (valence, arousal, dominance, liking). The labels are continuous values in the range 1.0-9.0. We make the labels discrete using the Emotion model as discussed next.

TABLE 1. Details of the DEAP pre-processed dataset.

Subjects	Recordings/ subject	EEG channels/ recording	Samples/ channel	Ratings/ trial
32	40	40	8064	valence,arousal dominance, liking

B. EMOTION RECOGNITION SETTINGS

Emotions represent mental and physiological states of human beings in terms of diverse types of thoughts, feelings and behaviors. A number of theories in cognition, psychology and neuroscience exist for the study and analysis of human emotions. However, it is challenging to determine how are emotions defined and differentiated [50].

In literature, there are two models for the theoretical emotion representation. These are called discrete emotion model (DEM) [51] and the bi-dimensional emotion model (BEM) [52]. In the DEM, a variety of emotions including fear, anger, happiness, sadness,surprise and disgust are defined as basic human emotions [51]. The BEM model embodies different emotional states on a multidimensional scale represented by arousal and valence basis. The BEM model is mostly used in the emotion recognition literature due to simplicity and generality. The BEM model categorizes emotions into four groups in the valence arousal space: low arousal/low valence (LALV), low arousal/high valence (LAHV), high arousal/low valence (HALV) and high arousal/high valence (HAHV). In our emotion recognition experiments, we used the BEM model which is based on the valence arousal classification. In the DEAP dataset each dimension is represented by values in the range 1 to 9. For two class classification, the labels are decided using the rating value less than 5 and greater than 5. Similarly, for four class classification the valence arousal space can be divided into four quadrants (LALV, HALV, LAHV and HAHV) according to the ratings.

C. EXPERIMENTAL SETUP

We use the leave-one-subject-out strategy for evaluating our proposed algorithm. This is done by training the model on 31 subjects and testing on the remaining one subject. The experiment is repeated 32 times with different training and testing data configurations and the average accuracy is reported. For computing the base kernel representations, the best value for parameter σ in (Eq. 1) is found automatically from only the available training kernels by employing the binary search method of Lin *et al.* [53]. We set the parameter C of SVM to a small value of 100. The target sparsity value s in NNSPCA algorithm is experimentally set in the range $\{1, 3, 5, 7, 9\}$.

VI. RESULTS

A. PERFORMANCE OF INDIVIDUAL CHANNELS

We first evaluate the accuracy of individual channels for emotion recognition in two class (valance, arousal) classification problem. We use our proposed feature extraction methods to represent each channel and then use SVM classifier for individual channels. The average accuracy achieved by each channel in 32 experiments are shown in Fig. 3. It can be observed that some of the channels have low accuracy due to their limited discrimination ability.

In the next experiment, we evaluate the accuracy of using multiple channels to learn individual classifiers and then integrate their performances at decision level. We use different classifier combination strategies such as sum based score fusion, majority voting [54] and kernel averaging. For this purpose, we train one K-NN classifier per channel using the proposed feature representations. We then combine the results at score level using sum rule of score fusion. We set $K = 3$ in these experiments. Next, we learn one SVM classifier per channel and fuse their results the decision level using majority voting scheme. Similarly, the performance of channel level fusion by concatenating feature representation of all the channels and learn the resulting feature vector for classification. Finally, we compare the results of the proposed method with the performance of a kernel SVM with an average kernel of all the channels. Table 2 summarizes the results of these experiments. The proposed SDEL algorithm performs better than the simple classifier combination strategies due its discriminative channel selection and combination. Moreover, we only train one classifier (K-SVM) for all the channels in an efficient manner.

TABLE 2. Average classification accuracy (%) in 2-class classification setting.

	Arousal	Valence
Ensemble method		
K-NN (sum score fusion)	60.4	55.4
SVM (majority voting)	63.1	59.1
SVM (channel fusion)	62.3	56.2
K-SVM (average kernel)	66.7	64.7
Proposed SDEL	**70.1**	**77.4**

B. COMPARISON WITH OTHER METHODS

we also compare the results of the proposed algorithm with other recent EEG based emotion recognition algorithms. These algorithms include the Bayesian classifier based method [55], the DEAP method [8], Ontology based method [56], Segment Level Decision Fusion (SLDF) [57], Sparsity constrained differential evolution (SCDE) based channel selection [58] and Empirical Mode Decomposition (EMD) [58]. Table 3 summarizes our comparison with the existing method on DEAP dataset. Due to the discriminative kernel learning, the proposed algorithm has outperformed the existing methods in the two class classification experiments. The Ontology based method [56], SCDE [58] and EMD [59] methods have better accuracy for the Arousal class. However these methods only use the data of 8 or 20 subjects in their experiments while we use the data of all 32 subjects.

C. SDEL AS A PREPROCESSING STAGE FOR DISCRIMINATIVE CHANNEL SELECTION

One application of our proposed algorithm is discriminative channel selection. We evaluate this capability using our algorithm as a preprocessing stage to improve a current deep learning based EEG emotion recognition algorithm presented

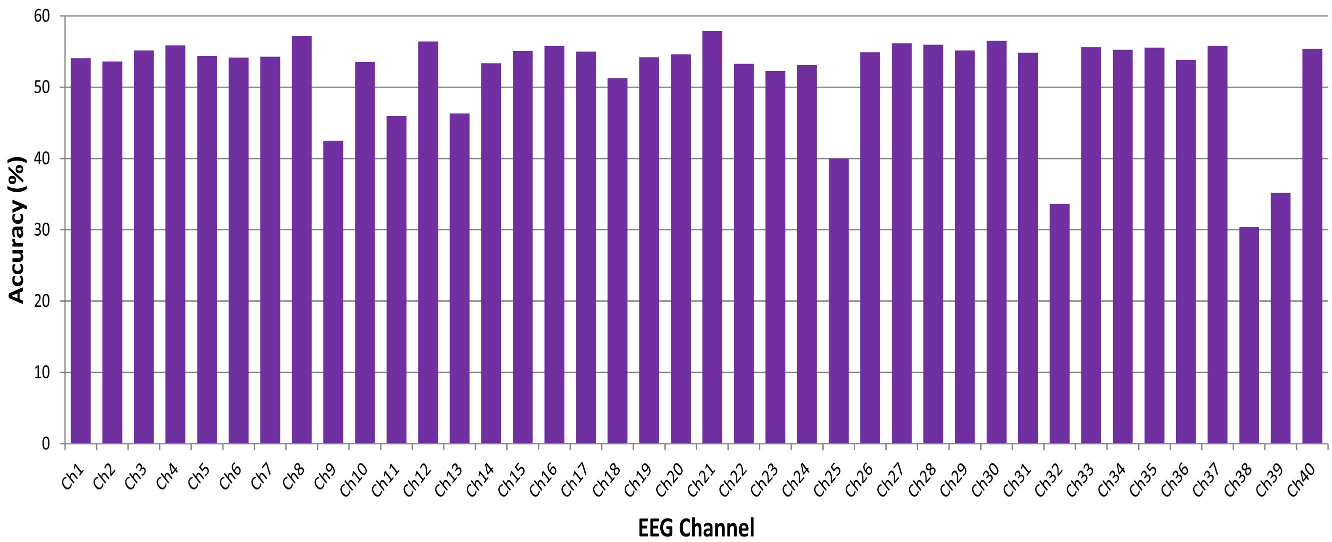


FIGURE 3. Average accuracy of individual channels in DEAP dataset using SVM classifier for two class emotion classification.

TABLE 3. Average classification accuracy (%) in 2-class classification setting.

Method	Arousal	Valence
Bayesian [55]	66.4	66.6
DEAP [8]	62.0	57.6
Ontology Based [56] (only 8 subjects)	75.1	81.7
SLDF [57]	68.4	76.9
SCDE Channel Selection [58] (only 8 subjects)	74.5	73.0
EMD [59] (only 20 subjects)	75.0	72.8
Proposed SDEL	**70.1**	**77.4**

TABLE 4. Average classification accuracy (%) in 2-class classification setting. The proposed algorithm improves the accuracy by using only the most discriminative channels.

	2 classes			
	Arousal		Valence	
	Without SDEL	**With SDEL**	Without SDEL	**With SDEL**
DNN [33]	73.12	**73.51**	75.78	**77.27**
CNN [33]	73.36	**74.53**	81.40	**82.81**

by Tripathi *et al.* [33]. They trained different models of deep and convolutional neural networks on DEAP dataset and reported excellent results on emotion recognition for both two class valence, arousal (low, high) as well as three class valence, arousal (high, normal, low) classification. In our experiments, we first identify the most discriminative channels using our SDEL algorithm (cannels having non-zero weights). Next, we use only our selected channels to train the best performing deep model of [33] and then record the accuracy. We use the same features along with our features to represent each channel as done in [33]. Table 4 and Table 5 summarize the results of using the proposed SDEL with the deep learning based methods. Our algorithm uses less than

TABLE 5. Average classification accuracy (%) in 3-class classification setting. The proposed algorithm improves the accuracy by using only the most discriminative channels.

	3 classes			
	Arousal		Valence	
	Without SDEL	**With SDEL**	Without SDEL	**With SDEL**
DNN [33]	55.70	**55.93**	58.44	**59.06**
CNN [33]	57.58	**60.23**	66.79	**68.20**

40 channels on the average and has increased the accuracy of the deep models. This is due to the high quality and more discriminative selected channels which help learning better supervised deep models and reduce over-fitting.

VII. COMPUTATIONAL TIME

The overall computational time of the proposed method include the time for computing the feature representations, the kernel representations, the scatter matrices and the time taken by the NNSPCA algorithm. During the training stage these operations are performed off-line. As shown in [42], the complexity of the NNSPCA is near-linear time. The test execution time of the proposed algorithm is fast because features and kernel representations of only the selected channels needs to be computed. Specifically, on Intel Corei7 3.8GHz CPU with 32GB RAM and MATLAB implementation, on the DEAP dataset, the time taken for computing the features and kernels in the training stage was more than 1500 seconds while the time for NNSPCA was 0.1 seconds. The testing time for classifying one EEG data matrix $\mathbf{R}_j$ was 2.01 seconds.

VIII. CONCLUSION

We proposed an ensemble learning algorithm for automatic EEG channel selection and combination for the application of emotion classification. We represented channels using kernels and formulated ensemble learning as graph embedding linear discriminant analysis objective function which was solved using a sparse non-negative principal component analysis algorithm. Due to sparse learning, only the most discriminative channels were included in the final ensemble. Experiments on standard EEG datasets for emotion recognition verified that the proposed method significantly improves the performance of emotion recognition.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank . . . ." Instead, write "F. A. Author thanks . . . ." In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

REFERENCES

[1] A. Etkin, C. Büchel, and J. J. Gross, "The neural bases of emotion regulation," *Nature Rev. Neurosci.*, vol. 16, no. 11, pp. 693–700, 2015.
[2] B. Cheng and G. Liu, "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *Proc. 2nd Int. Conf. Bioinf. Biomed. Eng. (ICBBE)*, 2008, pp. 1363–1366.
[3] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 102–115, Jan./Mar. 2012.
[4] I. Fernández-Varela, E. Hernández-Pereira, D. Álvarez-Estévez, and V. Moret-Bonillo, "Combining machine learning models for the automatic detection of EEG arousals," *Neurocomputing*, vol. 268, pp. 100–108, Dec. 2017.
[5] R. W. Picard, *Affective Computing*, vol. 252. Cambridge, MA, USA: MIT Press, 1997.
[6] EMOTIV. (Dec. 2017). *Eeg Device EPOC*. [Online]. Available: https://www.emotiv.com/epoc/
[7] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, to be published.
[8] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Oct./Mar. 2012.
[9] Y. Zhang, X. Ji, B. Liu, D. Huang, F. Xie, and Y. Zhang, "Combined feature extraction method for classification of EEG signals," *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3153–3161, Nov. 2017.
[10] B. Hu, X. Li, S. Sun, and M. Ratcliffe, "Attention recognition in EEG-based affective learning research using CFS+KNN algorithm," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 38–45, Jan./Feb. 2016.
[11] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges," *Brain-Comput. Interfaces.*, vol. 1, no. 2, pp. 66–84, 2014.
[12] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017.
[13] J. Zhang, Z. Yin, and R. Wang, "Pattern classification of instantaneous cognitive task-load through GMM clustering, laplacian eigenmap, and ensemble SVMs," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 947–965, Jul./Aug. 2017.
[14] T. T. Erguzel, G. H. Sayar, and N. Tarhan, "Artificial intelligence approach to classify unipolar and bipolar depressive disorders," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1607–1616, Aug. 2016.
[15] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.
[16] J. S. Ebersole and T. A. Pedley, *Current Practice of Clinical Electroencephalography*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2003.
[17] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, 2015.
[18] Z. Zhang, F. Li, T. W. S. Chow, L. Zhang, and S. Yan, "Sparse codes auto-extractor for classification: A joint embedding and dictionary learning framework for representation," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3790–3805, Jul. 2016.
[19] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image Vis. Comput.*, vol. 32, no. 10, pp. 814–823, 2014.
[20] Z. Zhang *et al.*, "Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3798–3814, Aug. 2018.
[21] K. Li, Z. Ding, S. Li, and Y. Fu, "Discriminative semi-coupled projective dictionary learning for low-resolution person re-identification," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018.
[22] Z. Zhang, F. Li, M. Zhao, L. Zhang, and S. Yan, "Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2429–2443, Jun. 2016.
[23] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from EEG signals using multidimensional information in EMD domain," *BioMed Res. Int.*, vol. 2017, 2017.
[24] R. Khosrowabadi, C. Quek, K. K. Ang, and A. Wahab, "ERNN: A biologically inspired feedforward neural network to discriminate emotion from EEG signal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 609–620, Mar. 2014.
[25] M. Sreeshakthy and J. Preethi, "Classification of emotion from EEG using hybrid radial basis function networks with elitist PSO," in *Proc. IEEE 9th Int. Conf. Intell. Syst. Control (ISCO)*, Jan. 2015, pp. 1–4.
[26] M. Murugappan, "Human emotion classification using wavelet transform and KNN," in *Proc. Int. Conf. Pattern Anal. Intell. Robot. (ICPAIR)*, vol. 1, Jun. 2011, pp. 148–153.
[27] A. Jalilifard, E. B. Pizzolato, and M. K. Islam, "Emotion classification using single-channel scalp-EEG recording," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 845–849.
[28] M. Murugappan, R. Nagarajan, and S. Yaacob, "Combining spatial filtering and wavelet transform for classifying human emotions using eeg signals," *J. Med. Biol. Eng.*, vol. 31, no. 1, pp. 45–51, 2011.
[29] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017.
[30] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan./Mar. 2016.
[31] Y. Gao, H. J. Lee, and R. M. Mehmood, "Deep learninig of EEG signals for emotion recognition," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun./Jul. 2015, pp. 1–5.
[32] F. Weninger, F. Eyben, and B. Schuller, "On-line continuous-time music mood regression with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 5412–5416.
[33] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proc. AAAI*, 2017, pp. 4746–4752.
[34] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affective Comput.*, to be published.
[35] A. Al-Ani and M. Mesbah, "EEG rhythm/channel selection for fuzzy rule-based alertness state characterization," *Neural Comput. Appl.*, vol. 30, no. 7, pp. 2257–2267, 2016.
[36] P. Ackermann, C. Kohlschein, J. Á. Bitsch, K. Wehrle, and S. Jeschke, "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods," in *Proc. IEEE 18th Int. Conf. E-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–6.
[37] N. Masood, H. Farooq, and I. Mustafa, "Selection of EEG channels based on spatial filter weights," in *Proc. IEEE Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Mar. 2017, pp. 341–345.
[38] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017.

[39] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 1241–1245.

[40] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-based EEG sensor selection methods for emotion recognition," *Sensors*, vol. 16, no. 10, p. 1558, 2016.

[41] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, 2007.

[42] M. Asteris, D. Papailiopoulos, and A. Dimakis, "Nonnegative sparse PCA with provable guarantees," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1728–1736.

[43] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[44] P. H. Chen, C. J. Lin, and B. Schölkopf, "A tutorial on ν-support vector machines," *Appl. Stochastic Models Bus. Ind.*, vol. 21, no. 2, pp. 111–136, 2005.

[45] F. Bao, X. Liu, and C. Zhang, "PyEEG: An open source python module for EEG/MEG feature extraction," *Comput. Intell. Neurosci.*, vol. 2011, 2011.

[46] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proc. 8th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 1995, pp. 212–217.

[47] C. J. James and D. Lowe, "Extracting multisource brain activity from a single electromagnetic channel," *Artif. Intell. Med.*, vol. 28, no. 1, pp. 89–104, 2003.

[48] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affective Comput.*, to be published.

[49] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. 100, no. 1, pp. 90–93, Jan. 1974.

[50] R. S. Lazarus, *Emotion and Adaptation*. London, U.K.: Oxford Univ. Press Demand, 1991.

[51] T. Dalgleish and M. Power, *Handbook of Cognition and Emotion*. Hoboken, NJ, USA: Wiley, 2000.

[52] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Develop. Psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[53] Y.-Y. Lin, T.-L. Liu, and C.-S. Fuh, "Multiple kernel learning for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1147–1160, Jun. 2011.

[54] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[55] S. Y. Chung and H. J. Yoon, "Affective classification using Bayesian classifier and supervised learning," in *Proc. 12th Int. Conf. Control, Automat. Syst. (ICCAS)*, Oct. 2012, pp. 1768–1771.

[56] X. Zhang, B. Hu, J. Chen, and P. Moore, "Ontology-based context modeling for emotion recognition in an intelligent Web," *World Wide Web*, vol. 16, no. 4, pp. 497–513, 2013.

[57] V. Rozgić, S. N. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 1286–1290.

[58] Y. Dai, X. Wang, P. Zhang, W. Zhang, and J. Chen, "Sparsity constrained differential evolution enabled feature-channel-sample hybrid selection for daily-life EEG emotion recognition," *Multimedia Tools Appl.*, vol. 77, no. 17, pp. 21967–21994, 2018.

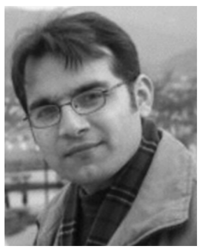
[59] A. Mert and A. Akan, "Emotion recognition from EEG signals by using multivariate empirical mode decomposition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 81–89, 2018.

HABIB ULLAH received the Ph.D. degree from the University of Trento, Italy, in 2015, and the M.Sc. degree in computer engineering from Hanyang University, Seoul, South Korea, in 2009. In 2015, he received a prestigious competitive Postdoctoral Fellowship from the École de Technologie Supérieure, Montreal, Canada. He is currently an Assistant Professor with the University of Ha'il, Saudi Arabia. His research interests include computer vision and machine learning.

MUHAMMAD UZAIR received the Ph.D. degree from the University of Western Australia and the M.S. degree from Hanyang University, South Korea. He was an Assistant Professor in electrical engineering with COMSATS University Islamabad, Wah Campus, Pakistan. His current research interests include computer vision, signal processing, and applied machine learning.

ARIF MAHMOOD received the B.E. degree in civil engineering from UET Lahore, Pakistan, in 1994, and the M.S. degree in computer science and the Ph.D. degree from the LUMS School of Science and Engineering, Lahore, Pakistan, in 2003 and 2011, respectively. He is currently an Associate Professor with Information Technology University, Pakistan. His research interests include image and video compression, object tracking, and image segmentation and classification.

MOHIB ULLAH received the bachelor's degree in electronic and computer engineering from the Politecnico di Torino, Italy, in 2012, and the master's degree in telecommunication engineering from the University of Trento, Italy, in 2015. He is currently pursuing the Ph.D. degree in computer science from the Norwegian University of Science and Technology, Norway. His research interests include crowd analysis, object segmentation, behavior classification, and tracking.



SULTAN DAUD KHAN received the M.Sc. degree in electronics and communication engineering from Hanyang University, South Korea, and the Ph.D. degree from the University of Milano-Bicocca, Italy. His research interests include computer vision application to pedestrian and crowd analysis.

FAOUZI ALAYA CHEIKH (SM) received the Ph.D. degree in information technology from the Tampere University of Technology, Tampere, Finland, in 2004. He has been a Researcher with the Signal Processing Algorithm Group, since 1994. Since 2006, he has been affiliated with the Department of Computer Science and Media Technology, Gjøvik University College, Norway, as an Associate Professor. Since 2016, he has been with the Norwegian University of Science and Technology. He teaches courses on image and video processing and analysis and media security. His research interests include e-learning, 3D imaging, image and video processing and analysis, video-based navigation, biometrics, pattern recognition, embedded systems, and content-based image retrieval. In these areas, he has published more than 100 peer-reviewed journal and conference papers, and supervised four post-doc researchers, five Ph.D. and a number of M.Sc. thesis projects. He is currently the Co-Supervisor of five Ph.D. students. He has been involved in several European and national projects among them: ESPRIT, NOBLESS, COST 211Quat, HyPerCept, IQ-Med, and H2020 ITN HiPerNav. He is a member of NOBIM and Forskerforbundet (The Norwegian Association of Researchers - NAR.) He is on the Editorial Board of the *IET Image Processing Journal* and the Editorial Board of the *Journal of Advanced Robotics and Automation* and the technical committees of several international conferences. He is an Expert Reviewer to a number of scientific journals and conferences related to the field of his research.