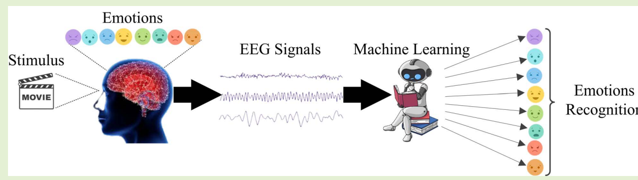
IEEE Sensors Council

Ensemble Machine Learning-Based Affective Computing for Emotion Recognition Using Dual-Decomposed EEG Signals

Kranti S. Kamble, *Graduate Student Member, IEEE*, and Joydeep Sengupta



***Abstract*—Machine learning (ML)-based algorithms have shown promising results in electroencephalogram (EEG)-based emotion recognition. This study compares five ensemble learning-based ML (EML) algorithms with five conventional ML (CML) algorithms for recognizing multiple human emotions from EEG signals. A publicly available DREAMER database having nine emotions is used to design ML-based system, which is validated on SEED, INTERFACES, and MUSEC databases. In this study, initially, EEG signals are separated into theta, alpha, beta, and gamma bands by applying discrete wavelet transform and then empirical mode decomposition is applied for further decomposition of band-separated EEG signals into intrinsic mode functions (IMFs). Then, 31 statistical features are extracted from IMFs to design ML-based system using five multiclass EML algorithms such as bagging, random forest, rotation random forest, extreme gradient boost, and adaptive boosting. Finally, the performance of these five EML algorithms is evaluated using 10-fold cross-validation and compared against five CML algorithms using performance evaluation metrics such as accuracy, F1-score, kappa-score, and area-under-the-curve (AUC). The mean accuracy of multiclass emotion recognition over five EML algorithms is ∼5.87% and ∼6.08% higher than the mean accuracy of five CML algorithms, for both arousal (88.95% *vs.* 83.08%) and valence (88.90% *vs.* 82.81%) dimensions, respectively. The EML-based bagging algorithm reported the highest accuracy, F1-score, kappa-score, and AUC of 95.81%, 0.81, 0.79, and 0.98, respectively for arousal and 95.53%, 0.80, 0.77, and 0.98, respectively for valence. A similar trend is also observed on the three validation datasets. The EML algorithms provide better multiclass emotion recognition compared to CML algorithms.**

***Index Terms*—EEG, emotion recognition, multiclass classification, ensemble learning, machine learning, DREAMER.**

I. INTRODUCTION

HUMAN emotion recognition has become one of the essential moves towards advanced human-machine interactions. Emotions are associated with the psychological and physiological conditions of cognition and awareness [1]. EEG signals are controlled by the autonomic nervous system of a person and are not influenced by a person's subjective aspects [2]. Moreover, EEG signals are reliable in expressing human emotion recognition, because they directly capture electrical activities from the human's brain cortex and provide true information about several mental states [3].

In general, emotions are categorized into two types: (i) discrete emotions [4] such as surprise, fear, joy, sadness, anger, and disgust, and (ii) multi-dimensions emotions such as arousal and valence dimensions [5], [6]. The arousal dimension explains the thrilled or passive nature of emotions, whereas valence indicates the degree to which the emotion is positive or negative [5], [6]. In this study, we characterized human emotions in arousal and valence dimensions.

In the past few years, artificial intelligence (AI)-based machine learning (ML) and deep learning (DL) algorithms are consistently being used in EEG-based emotion recognition [7]–[10]. Such AI-based algorithms generally follow a two-tier framework: (i) feature extraction from EEG signals and (ii) classification of these features into human emotions using ML or DL algorithms. In the first tier, initially, the EEG signal is decomposed into frequency bands (alpha, beta, gamma, and theta) using several methods such as band-pass filtering, short-time Fourier transform (STFT), discrete wavelet transform (DWT), flexible analytic wavelet transform (FAWT), tunable-Q wavelet transforms (TQWT), empirical mode decomposition (EMD), and empirical wavelet transform (EWT) [11]–[16]. Then from each frequency band, descriptive features are extracted. In the second tier, descriptive features from band-separated EEG signals are transformed into several mental emotion states using AI-based algorithms.

Manuscript received November 29, 2021; accepted December 12, 2021. Date of publication December 15, 2021; date of current version January 31, 2022. This work was supported by the Ministry of Education, Government of India, to conduct this research work as a part of a Ph.D. dissertation at the Visvesvaraya National Institute of Technology, Nagpur, India. The associate editor coordinating the review of this article and approving it for publication was Prof. Shih-Chia Huang. *(Corresponding author: Kranti S. Kamble.)*

The authors are with the Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology Nagpur, Nagpur 440010, India (e-mail: kmble.kranti@gmail.com; joydeep.sengupta44@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2021.3135953, provided by the authors.
Digital Object Identifier 10.1109/JSEN.2021.3135953

Recent studies have followed this two-tier framework for EEG-based emotion recognition [17]–[20]. Recently, Tuncer *et al.* [17] proposed an automated EEG-based emotion recognition system using the support vector machine (SVM) algorithm, with an overall accuracy of 99.82%. Their system extracted the fractal features from the EEG signals, which were decomposed using the TQWT method. Lakhan *et al.* [18] also proposed EEG-based emotion recognition system to identify the arousal and valence dimensions using the SVM algorithm. Authors extracted power spectral density features from band-separated EEG signals. The overall classification accuracy for arousal and valence emotions were 67.4% and 66.67%, respectively. Sharma and Bhattacharyya [20] used the sliding mode singular spectrum analysis method for EEG decomposition and used the KNN algorithm to predict the human emotional states. The authors reported an overall classification accuracy of 92.38%. Similarly, other studies have also been published that used ML-based algorithms for emotion recognition [10]–[12], [21]–[24].

A common observation from these studies is that nearly all the studies have used conventional ML-based algorithms such as SVM or KNN, which has resulted in low emotion recognition accuracy. The performance of any ML-based system is depending upon the type of algorithm used for emotion recognition. Therefore, to boost the overall emotion recognition accuracy, there is a need to explore more ML-based algorithms. Ensemble learning is an advanced concept that helps in boosting the overall accuracy of the ML-based system [25]–[27]. The dependability of the ensemble classifiers on the combination of dissimilar classifiers has achieved the best performance by eliminating each "oversight" in a single classifier [28]. Ensemble learning algorithms are based on the voting of several conventional weak ML classifiers [29] and provide better emotion recognition.

Recently, Gupta *et al.* [19] used an ensemble learning-based random forest (RF) classifier for emotions recognition using the DEAP and SEED datasets. The authors used the FAWT method for decomposition of EEG signals and reported overall classification accuracy of 79.95% and 79.99% for arousal and valence dimensions, respectively. Using the same ensemble learning-based algorithm, authors reported average emotion recognition accuracy of 83.33% on SEED dataset. Sangnark *et al.* [10] also used the RF algorithm for emotion recognition by revealing music preferences and indicated the highest emotion recognition accuracy of 84.64±1.59%. Another recent study by Subasi *et al.* [30] used a rotation forest ensemble algorithm to predict emotions in the SEED dataset. The authors used the TQWT method for EEG decomposition and feature extraction and reported an overall emotion recognition accuracy of 93%.

Motivated by these results, our presented study attempted to design an ML-based emotion recognition system using the ensemble ML learning (EML) algorithms and compared its performance against conventional ML-based (CML) algorithms.

Furthermore, from the literature, it has been observed that most of the studies used a single popular method such as EMD for EEG signals decomposition and then extracted features from the decomposed EEG signal for emotion recognition [31], [32]. Using EMD alone for signal decomposition is associated with two challenges [33], [34] such as (i) at low frequencies, EMD provides undesirable intrinsic mode functions (IMFs) which cannot provide a clear understanding about the EEG signal and (ii) the first IMF contains a wide range of frequencies, and, therefore, an appropriate decomposition into uniform frequency components is not possible. Therefore, to get an appropriate decomposition and better understanding of EEG signals, it is important to separate the EEG signal into narrow frequency bands using the popular DWT method [33], [34]. DWT converts the multiband EEG signal into a sequence of narrow-band signals and, therefore, supports the EMD method to get IMFs with more concentrated frequency components [33], [34]. A similar kind of dual decomposition approach is recently proposed by Ji *et al.* [34] for motor imagery recognition in BCI. Das and Bhuiyan [31], also used dual decomposition, however, the authors used EMD followed DWT for the decomposition of EEG signals, which may be associated with the above-mentioned two challenges. Besides these studies, dual decomposition is also explored by several recent studies [35], [36] emotion recognition and seizure detection applications.

These studies have led the foundation for the focus of our study to use the dual-decomposition approach using the popular DWT and EMD methods for emotion recognition. This is the first study that uses such type of dual-decomposition method under ensemble learning framework to perform emotion recognition on the DREAMER database, which is the OpenBCI database collected using the consumer-grade EEG measuring sensors.

The objective of this study is to design an ML-based system that can predict multiclass emotions from EEG signals using a concept of dual decomposition under the framework of ensemble learning. We used two types of ML-based algorithms: (i) CML and (ii) EML. Fig. 1 shows the global view of the proposed multiclass emotion recognition system. The overall multiclass emotion recognition system is divided into three stages: (i) dual decomposition, (ii) feature extraction, and (iii) multiclass emotion classification. We hypothesize that the ensemble ML algorithms can provide a better prediction of multiclass emotions compared to the conventional ML algorithms. The novelties of this study are:

- This is the first study that used the dual decomposition concept for decomposing EEG signals collected from the emotion recognition DREAMER database.
- The multiclass affective computing for emotion recognition is performed using a series of ensemble learning algorithms and performance is compared against the conventional ML-based algorithms.
- Comparative analysis is provided between five ensemble learning-based ML algorithms and five conventional ML algorithms.
- The performance of the proposed system is also verified on the three types of external databases such as SEED, INTERFACES, and MUSEC for binary and multiclass emotional states recognition.

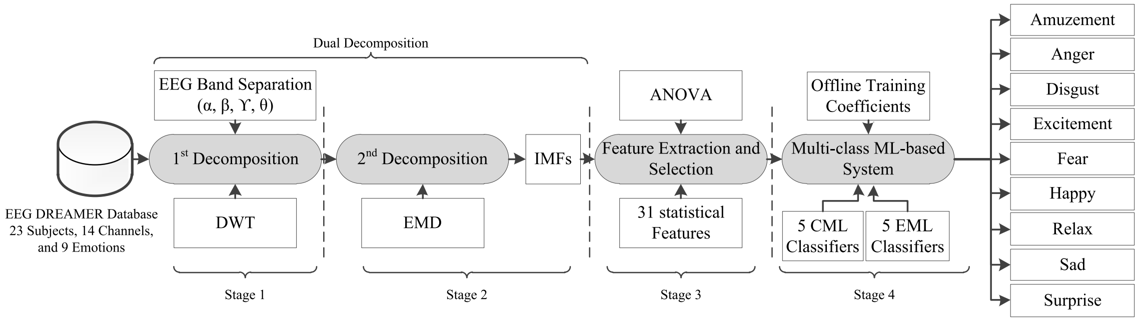


Fig. 1. Global view of the proposed ML-based multiclass emotion recognition system. DWT: discrete wavelet transform, EMD: empirical mode decomposition, IMF: intrinsic mode function, ANOVA: analysis of variance, CML: conventional machine learning, EML: ensemble machine learning.

II. Methods

A. Database

This study uses a multimodal DREAMER database that contains 14 channels of EEG signals recorded from 23 subjects (14 male, 9 female), aged between 22 and 33 years old (M = 26.6, SD = 2.7) by giving audio-visual stimuli in the form of 18 film clips [37]. The 18 film clips were divided into nine emotions such as amusement, anger, disgust, excitement, fear, happiness, calmness, sadness, and surprise. Each film clip continued for a period of 65 to 393 seconds, which is noted to be adequate for giving rise to single emotions. All the signals were recorded using portable, wearable, wireless, low-cost, and off-the-shelf equipment. EEG signals were captured at a sampling rate of 128 Hz, applying 16 gold-plated contact-sensors AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1, and M2 using an Emotive EPOC system with International 10-20 system and SHIMMER wireless sensor. A detailed protocol for this database is discussed in [37].

B. Dual Decomposition With DWT-EMD Combination

This study used DWT and EMD independently for decomposing input EEG signal and, therefore, it is termed as "dual decomposition". A similar kind of approach of using two different methods consecutively for decomposition has been tried before in previous studies [31], [34]–[36].

Each EEG signal is decomposed into four levels using DWT with Daubechies 5 (db5) wavelet. The DWT of the input EEG signal x(t) is mathematically expressed in Eq. (1).

$$w_x(j,k) = \int x(t)\psi^*_{j,k}(t)dt \tag{1}$$

where, $\psi_{j,k}(t)$ is a discrete wavelet basis function, which is expressed in Eq.(2) as,

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t-k), \quad j\epsilon Z,\ k\epsilon Z \tag{2}$$

and j and k are the scale and translation parameters. The four-level DWT has resulted in a series of approximations ($c_{i,j}$) and detailed coefficients ($d_{i,j}$) which are mathematically expressed in Eq.(3) and (4).

$$c_{j,k} = <x(t), \phi_{j,k}(t)>= \int x(t)2^{-j/2}\overline{\phi(2^{-j}t-k)}dt \tag{3}$$

$$d_{j,k} = <x(t), \psi_{j,k}(t)>= \int x(t)2^{-j/2}\overline{\psi(2^{-j}t-k)}dt \tag{4}$$

where, $\phi_{j,k}(t)$ is the basic scaling function and given as $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t-k)$. From the four-level decomposition, approximation and detailed coefficients corresponding to the four frequency bands called theta (4-7 Hz), alpha (8-13Hz), beta (14-30 Hz), and gamma (31-50 Hz) are extracted [38] and used to reconstruct the band-wise EEG signal. The generalized mathematical expression for the band-wise reconstruction of EEG signal by applying inverse DWT is given in Eq. (5)

$$x(t) = \sum_k c_{j,k}\phi_{j,k}(t) + \sum_k d_{j,k}\psi_{j,k}(t) \tag{5}$$

Alpha and theta bands show the valence dimension nature of the emotion, whereas beta and gamma show the arousal dimension of the emotions [39]–[41]. Stage 1 of Fig. 1 shows the band separation using DWT.

EMD is performed on the reconstructed band-wise separated EEG signals and decomposes the signal into multiple amplitudes and frequency modulated oscillatory patterns called intrinsic mode functions (IMFs) [13]. EMD has been implemented for rhythmic analysis and studying the instantaneous amplitude (IA) and instantaneous frequency (IF) components. Similar work of analyzing IA and IF of EEG signals also exists in [42], [43].

These IMFs satisfy two main conditions: (i) the number of extrema must be equal or differ by at most one from the number of zero crossings and (ii) at any point, the mean envelope value defined by local maxima and minima must be equal to zero. The first condition enhances the stationary conditions, while the second condition enhances the definition of instantaneous frequency (IF). For reconstructed band-wise separated EEG signals, x(t), the steps to calculate IMFs are as follows:

Step 1: Retain x(t) to the new variable $x_{old}(t)$ as $x_{old}(t) = x(t)$

Step 2: Find the local maxima (x_{max}) and minima (x_{min}) of $x_{old}(t)$.

Step 3: Generate envelopes e_{max} and e_{min} of maxima and minima using cubic spline interpolation.

Step 4: Compute mean envelope as

$$e_{mean}(t) = (e_{max} + e_{min})/2 \tag{6}$$

Step 5: Subtract the mean envelope of $x_{old}(t)$ from the original input x(t) to obtain the IMF. This is mathematically expressed as

$$x_{new}(t) = x_{old}(t) - e_{mean}(t) \tag{7}$$

Step 6: $x_{new}(t)$ is considered as IMF, if the two required IMF conditions are satisfied, else Steps 1-5 are repeated by assigning

$$x_{old}(t) = x_{new}(t) \tag{8}$$

Step 7: A new residue r is generated as:

$$r(t) = x_{old}(t) - c_i(t) \tag{9}$$

Steps 1 to 6 applied to the residue r, in order to find the remaining IMFs. The process continues till no further IMFs generate from the residue, and the last residue will be termed as r_e.

$$x(t) = r_e + \sum_{n=1}^{N} c_i(t) \tag{10}$$

where, N shows the number of IMFs extracted from the signal.

The minimum number of IMFs generated for all reconstructed EEG signals is six and, therefore, to ease the further analysis, this study decomposes all reconstructed EEG signals up to six IMFs, only. Stage 2 of Fig. 1 shows the decomposition of the reconstructed version of band-wise separated EEG signals using EMD. Each IMF will then be used to extract the features to train the multiclass ML-based algorithm. The overall sample size used for feature extraction is 69,552 (23 subjects $\times$ 14 channels $\times$ 9 emotions $\times$ 4 frequency bands $\times$ 6 IMFs).

C. Feature Extraction and Selection

In total, 31 types of statistical features are extracted from all 69,552 samples of the EEG signals. Feature extraction is performed separately for arousal and valence dimensions, each of 34776 samples. Table I has three columns, indicating the rank, features, and p-value of each of the 31 features. Feature selection is performed using the analysis-of-variance (ANOVA) method. The ANOVA method identifies dominant features that show significant association with the multiclass ground truth variable (in our case it is nine emotions). The ANOVA method ranks the feature based on the p-value. Out of 31 features, 26 features show a significant association with nine emotions. These 26 dominant features are then used to train the multiclass ML algorithms separately for arousal and valence dimensions. The ranking of these 26 features based p-values is shown in Table I. The feature extraction and selection is depicted in Stage 3 of Fig. 1. The mathematical formulae along with a brief discussion about these features have been provided in Table I from section I of the supplementary material.

D. Multiclass ML-Based Emotion Recognition System

Fig. 2 show local architecture of the multiclass emotion recognition system. The complete architecture is divided into two parts: (i) offline model and (ii) online model. The data partitioning block holds the ten-fold cross validation protocol that divides the input samples into training dataset and testing dataset. Offline model then uses the training dataset and extracts the 26 dominant statistical features using the ANOVA test. These 26 features along with the ground truth (i.e., nine emotions) are then used to train the offline multiclass ML-based algorithm. This study used two categories of ML-based algorithms: (i) five types of CML-based algorithms such as SVM [44], logistic regression (LR) [45], linear discriminant analysis (LDA) [46], [47], quadratic discriminant analysis (QDA) [47], and KNN [47] and (ii) five types of

TABLE I
RANKING OF FEATURES USING ANOVA

Rank	Features	p-value
1	Variance	$<$0.0001
2	Mean Absolute Deviation	$<$0.0001
3	Mean Absolute Value	$<$0.0001
4	Modified Mean Absolute Value	$<$0.0001
5	Modified Mean Absolute Value2	$<$0.0001
6	Root Mean Square	$<$0.0001
7	Standard Deviation	$<$0.0001
8	Inter Quartile Range	$<$0.0001
9	Median Absolute Deviation	$<$0.0001
10	Enhanced Mean Absolute Value	$<$0.0001
11	Log Detector	$<$0.0001
12	Willison Amplitude	$<$0.0001
13	Myopulse Percentage Rate	$<$0.0001
14	Entropy	$<$0.0001
15	Kurtosis	$<$0.0001
16	Minimum	$<$0.0001
17	Maximum	$<$0.0001
18	Maximum Fractal Length	$<$0.0001
19	Enhanced Wavelength	$<$0.0001
20	Difference Absolute STD Value	$<$0.0001
21	Log Energy Entropy	$<$0.0001
22	Average Amplitude Change	$<$0.0001
23	Zero Crossing	$<$0.0001
24	Wavelength	$<$0.0001
25	Slope Sign Change	$<$0.0001
26	Simple Square Integral	0.002
27	Skewness	0.361
28	Mean	0.442
29	Covariance	0.454
30	Median	0.860
31	Shanon Entropy	0.983

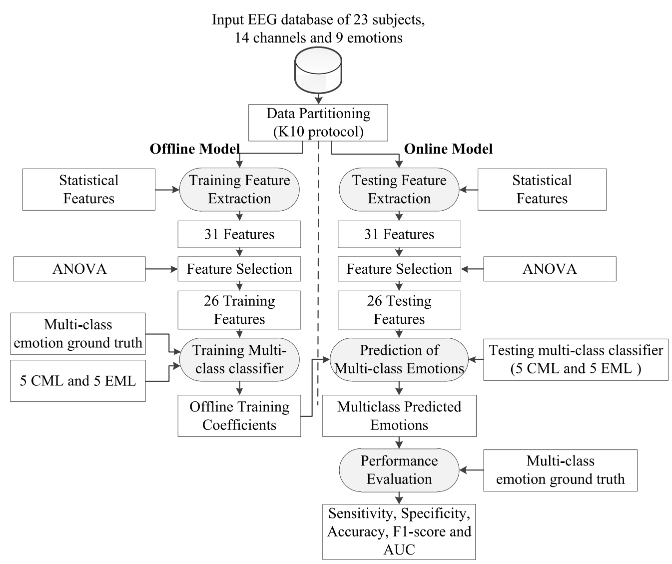


Fig. 2. Local architecture of the proposed multiclass ML-based emotion recognition system. CML: conventional machine learning, EML: ensemble machine learning, AUC: area under curve, ANOVA: analysis of variance.

TABLE II
GRID OF HYPERPARAMETERS FOR THE ML-BASED ALGORITHMS

Classifiers	Hyperparameters	Grid of Hyperparameters	Most probable optimized values Arousal	Valence
SVM	Kernel	['rbf','linear','poly']	rbf	rbf
	Gamma	[0.001, 0.01, 0.1, 1]	1	1
	C	[0.001, 0.01, 0.1, 1, 2]	2	2
	Degree	[1,2,3]	1	1
LR	penalty	[l1, l2]	l2	l2
	C	[1.0e-04, 7.7e-04, 5.9e-03, 4.6e-02, 3.5e-01, 2.8e+0, 2.2e+01, 1.7e+02, 1.3e+03, 1.0e+04]	0.01	1291.5
LDA	solver	['svd', 'lsqr', 'eigen']	svd	svd
QDA	reg_param	[0.1, 0.2, 0.3, 0.4, 0.5]	0.1	0.1
KNN	p	[1,2]	2	2
	n_neighbors	[1, 2, 3, 4, 5, 6, 7, 8, 9]	1	1
	leaf_size	[1, 2, 3, 4, 5, 6, 7, 8, 9]	2	7
Bagging	n_estimators	[100, 200, 500, 1000]	1000	1000
RF	n_estimators	[100, 200, 500, 1000]	500	200
	min_weight_fraction_leaf	[0.1, 0.01]	0.1	0.1
	min_samples_split	[2,5,10]	2	10
	min_samples_leaf	[2,5,10]	2	5
	max_leaf_nodes	[20, 40]	40	40
	max_depth	[3,4,5]	5	5
RFE	n_estimators	[10, 50, 70, 100, 125]	100	100
	K	[3, 5, 7]	5	5
AdaBoost	n_estimators	[100, 300, 500, 700, 900, 1000]	300	100
XGBoost	n_estimators	[100, 300, 500, 700, 1000]	300	300
	min_child_weight	[1, 3, 5, 7]	3	3
	max_depth	[3, 5, 7, 9]	9	9
	learning_rate	[0.01, 0.10]	0.1	0.1
	gamma	[0, 0.1, 0.2 , 0.3]	0.2	0.2
	colsample_bytree	[0.6, 0.7, 0.8, 0.9]	0.6	0.6

EML-based algorithms such as RF [25], rotation forest ensemble (RFE) [48], bagging [49], AdaBoost [50], and extreme gradient boosting (XGBoost) [26].

Once trained, the offline training coefficients are then used to transform the test features from the online testing dataset into the multiclass emotions using the testing multiclass classifier. The testing dataset is completely independent and is not part of the training dataset. The performance of the proposed multiclass ML-based emotion recognition system is evaluated using the six performance evaluation metrics such as F1-score, kappa-score, sensitivity, specificity, accuracy, and area-under-the-curve (AUC) and the brief discussion with their mathematical expression is provided in section III of the supplementary material. Note that the multiclass emotion ground truth was only used to evaluate the performance of the proposed system and is not part of the online testing model.

This study used nested cross-validation to train and evaluate the performance of the proposed ML-based system as well as to tune the hyperparameters of the ML classifiers. Nested cross-validation is the combination of two independent k-fold cross-validations with the inner loop is for tuning the hyperparameters and outer loop if for evaluating the model performance of an ML-based system [51], [52]. Therefore, in this study, we used 5-fold cross-validation for hyperparameter optimization using the randomsearchCV algorithm and 10-fold cross-validation for evaluating the performance of both EML and CML algorithms [51]. The randomsearchCV algorithm in python requires a grid of parameters as an input argument that contains the ranges for hyperparameters of the ML algorithms. Table II shows the grid and the most probable optimized values of hyperparameters for the ML-based algorithms. Note that the hyperparameters are optimized for each fold of 10-fold cross-validation and, therefore, there are 10 optimized values for each fold. Table III to Table X from section II the supplementary material shows the detailed hyperparameters for the DREAMER and other three validation databases.

E. Validation

1) SEED: The SJTU Emotion EEG Dataset (SEED) is a publicly available, 62 channel EEG database, collected from 15 subjects (7 males and 8 females; MEAN: 23.27, STD: 2.37). This database was collected in three sessions with an interval of one week. Each subject was given stimuli from 15 Chinese films with three targeted emotions (negative, positive, and neutral). These three targeted emotions were coded as -1 for negative (sad), 0 for neutral, and 1 for positive (happy). The sampling rate for this database was 200Hz. Out of 62 channels, Zheng and Lu [53] found that only 12 channels (FT7, FT8, T7, T8, C5, C6, TP7, TP8, CP5, CP6, P7, and P8) are informative and critically contribute to the emotion recognition. Therefore, in this study, we used the 12 critical channels for emotion recognition. Each of the 12 channel EEG signals is separated into four bands, each of which is separated into 4 IMFs. Therefore, the overall sample size used for feature extraction is 43,200 (15 subjects $\times$ 15 films, 12 channels $\times$ 4 frequency bands $\times$ 4 IMFs). Out of 31 statistical features, 25 dominant features (see Table II from section I of supplementary material) obtained using the ANOVA method are used to train the proposed ML-based system for multiclass emotion recognition.

2) INTERFACES: The second validation database, called INTERFACES, is an 8 channel (Fp1, Fp2, Fz, Cz, T3, T4, Pz, and Oz) EEG-based database, collected from 43 subjects (21 male and 22 female), aged between 16 and 34, using OpenBCI device [18]. The sampling rate for this database was 250Hz. In this database, audio-visual stimuli from 15 films were given to each subject to elicit emotions. Overall, there were 645 (43 subjects × 15 clips) samples. The targeted emotions were happiness, fear, and excitement. At the end of each film clip, the participants were asked to rate Valence (V), Arousal (A), Happiness (H), Fear (F), and Excitement (E) on a scale of 1–9. A sample with V or A score lower than 4.5 was labeled as low V or A, and vice versa. Out of 8 channels, authors reported dominant 4 channels (Fz, Cz, Pz, Oz) with better performance [18]. Therefore, the current study used these 4 channels for emotion recognition. Each EEG signal is separated into four bands, each of which is separated into 2 IMFs. Therefore, the overall sample size used for feature extraction is 20,640 (43 subjects × 15 films, 4 channels × 4 frequency bands × 2 IMFs). Feature extraction was performed separately for arousal and valence dimensions, each of 10320 samples.

3) MUSEC: This is a 62 channel, EEG-based music preference emotion recognition database, collected from 20 subjects (10 males, 10 females),aged between 22 and 35 (mean: 25.75, SD: 2.88) [10]. This database used the 110 popular Thai audio music clips (melody and song) as stimuli, which is divided into two binary classes as favored and non-favored. The sampling rate was 1200 Hz. Out of 62 channels, Sangnark *et al.* [10] found that 27 channels from the right hemisphere (Fp2, AF4-6-8, F2-4-6-8, FT8, FC2-4-6, T8, C2-4-6, TP8, CP2-4-6, P2-4-6-8, PO4-8, O2) and 11 channels from the right frontal (Fp2, AF4-8, F2-4-6-8, FT8, FC2-4-6) are informative and critically contribute for song and melody, respectively. So this current study used the same number of channels for emotion recognition for both melody and song groups, respectively. Initially, favored and non-favored emotions have 81 and 271 files for both melody and song groups, respectively. Therefore, the overall sample size used for feature extraction of melody is 12,001 ((77 favored + 271 non favored), 11 channels × 4 frequency bands × 1 IMF). The overall sample size used for feature extraction of a song is 29,349 (81 favored + 271 non favored), 27 channels × 4 frequency bands × 1 IMFs). The database was highly imbalanced in terms of class distribution. The nature of the ground truth label for this database is binary with favored and non-favored being the output classes for both song and melody stimuli. The same 31 statistical features are used to train the ML-based system. Note that in all the above three databases the minimum number of IMFs generated from all reconstructed EEG signals are considered in the analysis.

F. Statistical Analysis

All the statistical analysis was done in MATLAB2020b and using SPSS23.0. The feature selection is done using the ANOVA test with multiclass emotion ground truth. Dominant features are ranked using the 2-tailed p-value with the level of significance as 0.05. The entire multiclass emotion recognition system is built in an open-source python-based library called Scikit-learn [54].

III. Experiment Protocol

A. Selection of Optimal Classifier

In this experiment, an optimal ML-based classifier is selected from the 10 types of ML-based algorithms. All the

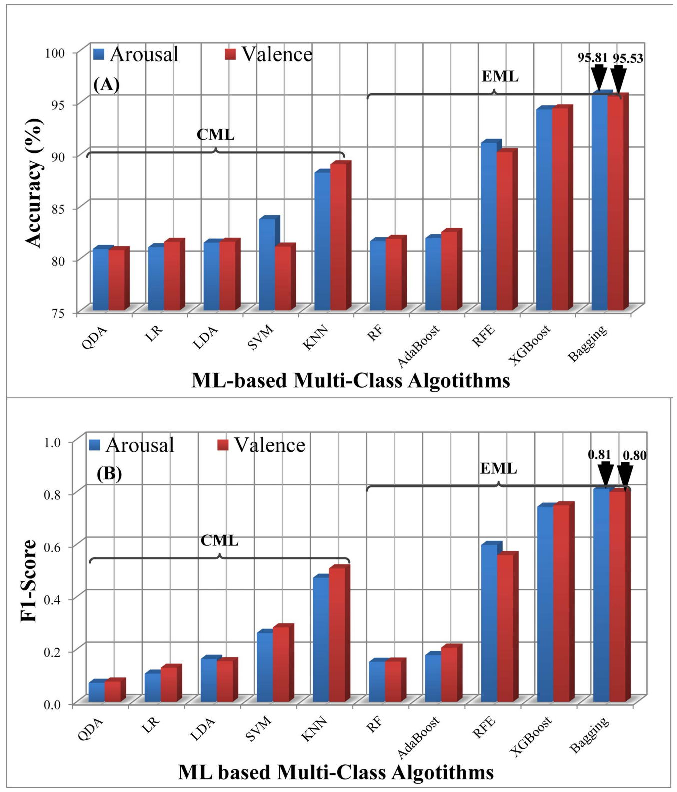


Fig. 3. Comparing the multiclass ML algorithms using (A) accuracy and (B) F1-score for both arousal and valence dimensions.

10 multiclass ML-based algorithms were trained using the 26 dominant features and multiclass emotion ground truth using 5 types of cross-validation protocols such as K2, K3, K4, K5, and K10. Then, the performance of these 10 ML-based algorithms is compared using the accuracy, F1-score, and AUC metrics. The ML-based algorithm with the highest values of these performance evaluation metrics is then selected as an optimal ML classifier. This optimal ML-based algorithm is then used to perform further analysis.

B. Selection of Best Cross-Validation (CV) Protocol

In this experiment, the performance of the proposed multiclass emotion recognition system is evaluated using accuracy and F1-score for five types of K-fold cross-validation protocols such as K2, K3, K4, K5, and K10. The cross-validation with the highest multiclass accuracy and F1-score for both arousal and valence dimension is then selected as the best cross-validation protocol and used for further analysis.

IV. Results

Fig. 3 compares the performance of CML and EML-based algorithms used in multiclass emotion recognition system. The bar charts in Fig. 3 compare the accuracy and F1-score for both arousal and valence dimensions.

As shown in Fig. 3 (A), for arousal dimension, the overall accuracies for multiclass emotion recognition for SVM, LR, LDA, QDA, KNN, RF, RFE, Bagging, AdaBoost and XGBoost are 83.75%, 81.06%, 81.5%, 80.89%, 88.22%,

TABLE III
COMPARING THE EML AND CML ALGORITHMS FOR MULTICLASS EMOTION RECOGNITION ON DREAMER DATABASE

Sr#	Type of Algorithm	Algorithm	Sensitivity (%)		Specificity (%)		Kappa-Score		F1-Score		AUC		Accuracy (%)	
			Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
1	CML	QDA	14.05	13.47	89.25	89.18	0.03	0.03	0.07	0.08	0.55	0.55	80.89	80.77
2		SVM	26.9	28.52	90.86	91.06	0.18	0.20	0.26	0.28	0.76	0.77	83.75	81.12
3		LR	14.79	17.09	89.35	89.63	0.02	0.04	0.11	0.13	0.64	0.68	81.06	81.57
4		LDA	16.77	17.14	89.6	89.64	0.07	0.07	0.16	0.15	0.64	0.66	81.5	81.58
5		KNN	47	50.6	93.38	93.83	0.41	0.45	0.47	0.51	0.76	0.72	88.22	89.02
		Average	23.90	25.36	90.49	90.67	0.14	0.16	0.22	0.23	0.67	0.68	83.08	82.81
6	EML	RF	17.36	18.38	89.67	89.8	0.06	0.07	0.15	0.15	0.64	0.64	81.63	81.86
7		AdaBoost	18.69	21.35	89.84	90.17	0.08	0.11	0.18	0.21	0.7	0.7	81.93	82.52
8		RFE	59.87	55.82	94.98	94.48	0.55	0.50	0.60	0.56	0.7	0.68	91.08	90.18
9		XGBoost	74.37	74.77	96.8	96.85	0.71	0.72	0.74	0.75	0.96	0.96	94.3	94.39
10		Bagging	81.14	79.87	97.64	97.48	0.79	0.77	0.81	0.80	0.98	0.98	95.81	95.53
		Average	50.29	50.04	93.79	93.76	0.44	0.43	0.50	0.49	0.79	0.79	88.95	88.90
Improvement for EML over CML(%)			**26.38**	**24.67**	**3.30**	**3.09**	**212.20**	**177.59**	**130.03**	**113.86**	**18.81**	**17.16**	**5.87**	**6.08**

81.63%, 91.08%, 95.81%, 81.93%, and 94.3%, respectively. Similarly, for valence dimension, these accuracies are 81.12%, 81.57%, 89.02%, 80.77%, 89.02%, 81.86%, 90.18%, 95.53%, 82.52%, and 94.39%, respectively for the corresponding ten ML algorithms. The mean accuracies over five CML algorithms are 83.08% and 82.81% for arousal and valence, respectively. The mean accuracies over five EML algorithms are 5.87% and 6.08% higher compared to the mean accuracies over five CML algorithms for arousal and valence, respectively. Fig. 3 (B), compares the F1-score for multiclass emotion recognition. As shown in Fig. 3 (B), for arousal dimension, the overall F1-scores for SVM, LR, LDA, QDA, KNN, RF, RFE, Bagging, AdaBoost and XGBoost are 0.26, 0.11, 0.16, 0.07, 0.47, 0.15, 0.60, 0.81, 0.18 and 0.74, respectively. Similarly, for valence dimension, these F1-scores are 0.28, 0.13, 0.15, 0.08, 0.51, 0.15, 0.56, 0.80, 0.21 and 0.75, respectively, for the corresponding ten ML algorithms. Out of five EML algorithms, Bagging reported the highest multiclass emotion recognition accuracy (95.81%), which is ~7.49% higher compared to the best conventional multiclass algorithm (i.e., KNN). Similarly, for valence, the multiclass emotion recognition accuracy of the Bagging algorithm (95.53%) was ~6.51% higher compared to the KNN algorithm. This shows that the ensemble algorithm can provide better emotion recognition compared to nearly all CML-based algorithms. Furthermore, Bagging is found to be the optimal ML algorithm to perform the emotion recognition on the DREAMER database. This was also true comparing the F1-score of the EML algorithm and the CML algorithms for both arousal and valence dimensions Fig. 3 (B). Table III provides the percentage improvement in the mean values of six performance metrics for both arousal and valence dimensions.

Fig. 4 compares the ROC plots for five EML algorithms with the ROC plots for the five CML algorithms. The mean AUCs over five CML algorithms are 0.67 and 0.68 for arousal and valence dimensions, respectively. Similarly, the mean AUCs over five EML algorithms is 0.79 for both arousal and valence dimensions, respectively. The mean AUCs over five EML algorithms are 18.81% (0.79 *vs.* 0.67, $p<0.0001$) and 17.16% (0.79 *vs.* 0.68, $p<0.0001$) higher compared to the mean AUCs over five CML algorithms for arousal and valence dimensions, respectively.

Fig. 4 also shows that the bagging algorithm is the best choice to perform the multiclass emotion recognition over the DREAMER database. From Fig. 4 (A) and Fig. 4 (B), the AUC for Bagging algorithm is ~28.2% (0.98 vs. 0.76, $p<0.0001$) and ~25.9% (0.98 vs. 0.77, $p<0.0001$) higher than the AUC values of the best CML algorithm in both arousal and valence dimensions, respectively. This also validates our hypothesis that EML algorithms can provide better emotion recognition compared to the CML algorithms. Fig. 5 (A) and Fig. 5 (B) demonstrates the effect of cross-validation protocols on the accuracy and F1-score, respectively. From Fig. 5, it is clear that the performance of the ML-based multiclass emotion recognition system improved with the cross-validation protocol. The K10 cross-validation protocol is found to be the best choice for evaluating the performance of the proposed emotion recognition system. These results are also in line with some previous studies that also reported better prediction using K10 protocol [55].

Fig. 6 compares the performance of the emotion recognition system in terms of accuracy and AUC for with and without feature selection using ANOVA. Without and with feature selection refers to the use of 31 original features and 26 dominant features, respectively. It has been observed that feature selection using ANOVA provides an improvement of ~2.31% (95.81% *vs.*93.50%) and ~1.25% (95.53% *vs.* 94.28%) in accuracy over without feature selection for both arousal and valence dimensions, respectively. Similarly, the improvement by using feature selection in terms of AUC is ~6.52% (0.98 *vs.* 0.92, $p<0.0001$) and ~2.08% (0.98 *vs.* 0.96) for both arousal and valence dimensions, respectively.

Fig. 4 (C) and (D) shows the emotion-wise ROC plots using the optimal classifier and the best cross-validation protocol for both arousal and valence dimensions. From Fig. 4 (C) and (D) it is clear that the bagging classifier has the high predictive ability for all nine types of emotions in both arousal and valence. Out of nine emotions, "sad" emotion has been recognized by the bagging classifier with highest AUC of 1 ($p<0.0001$) and 0.99 ($p<0.0001$) for arousal and valence dimensions, respectively. All nine emotions have high-predictive ability for bagging classifier with AUC$>$0.93. Table IV and Table V show the emotion-wise performance evaluation metrics of arousal and valence for the Bagging classifier respectively. The last column of Table IV and Table V

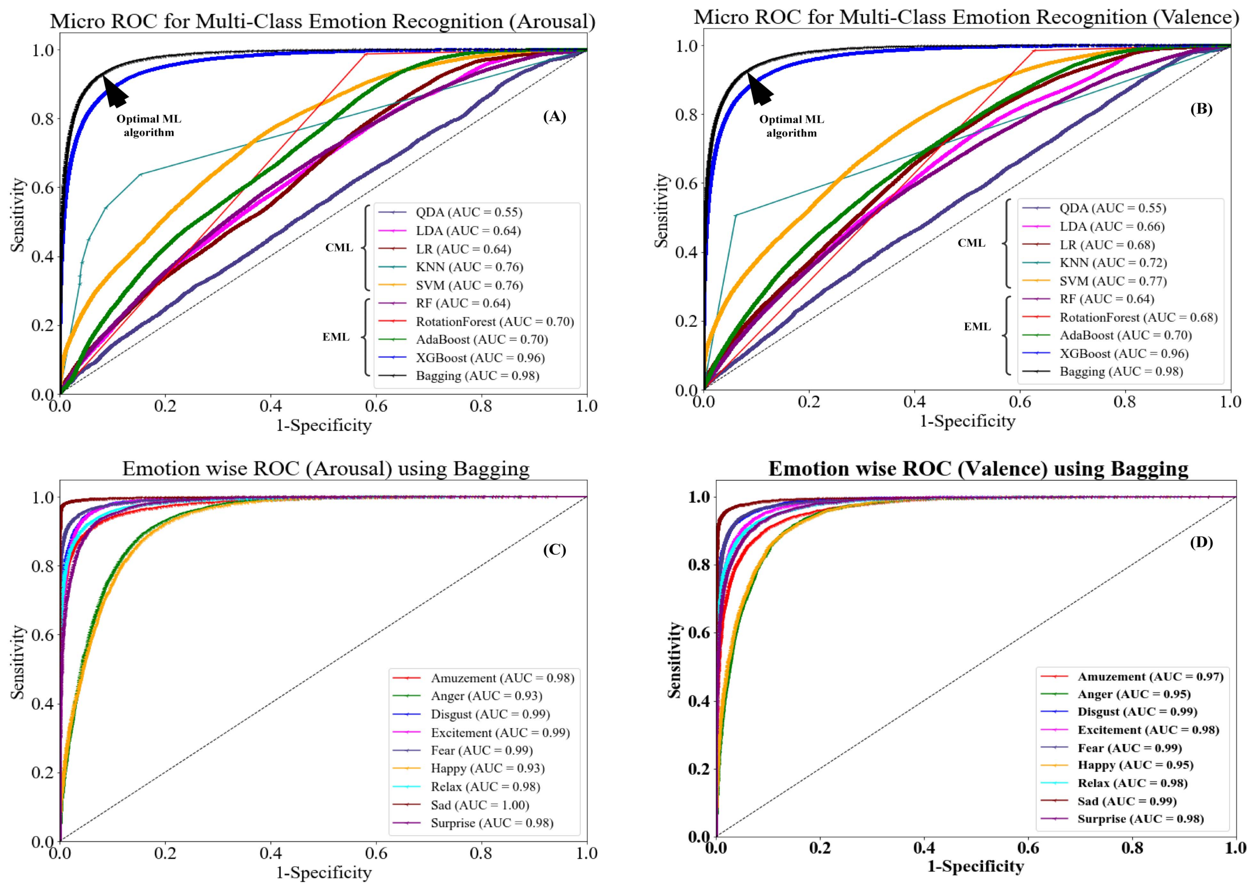


Fig. 4. ROC curves comparing the 2 categories of multiclass emotion recognition algorithms (EML *vs*. CML) for (A) arousal and (B) valence dimensions, Emotion-wise ROC curves for optimal ML based algorithm i.e., Bagging for (C) arousal and (D) valence dimensions.

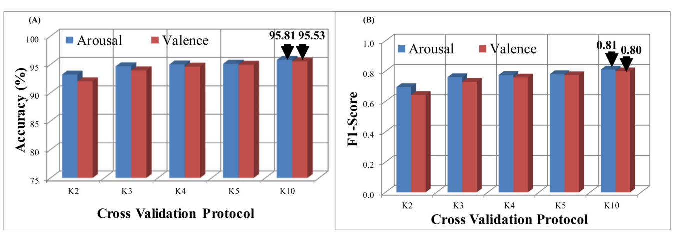


Fig. 5. Comparing the sensitivity of the multiclass emotion recognition system due to CV protocols using (A) Accuracy and (B) F1-score for both arousal and valence dimensions.

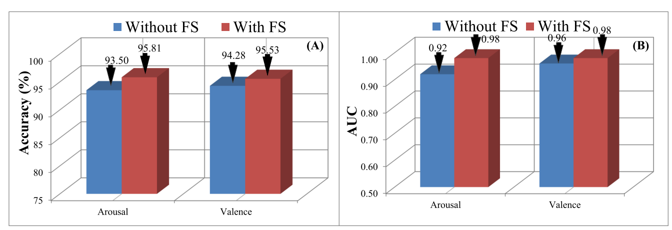


Fig. 6. Effect of feature selection (FS) on the performance of ML-based multiclass emotion recognition system.

shows the overall mean performance evaluation metrics over nine emotions.

Fig. 7 compares the performance of EML and CML algorithms in the proposed ML-based emotion recognition system for the SEED database. The bar charts in Fig. 7 (A) compare the average accuracies and in Fig. 7 (B) compare the average F1-scores over three sessions of the SEED database for both arousal and valence dimensions. The mean accuracies over five EML algorithms are 3.67% (81.39% *vs.* 77.73%) and 3.57% (79.71% *vs.* 76.4%) higher compared to the mean accuracies over five CML algorithms for arousal and valence dimensions, respectively Fig. 7 (A). Similarly, the mean F1-score over five EML algorithms are 9.82% (0.72 *vs.* 0.65) and 12.38% (0.69 *vs.* 0.61) higher compared to the mean F1-score over five CML algorithms for arousal and valence dimensions, respectively Fig. 7 (B).

The overall performance evaluation metrics for SEED is shown in VI. Note that on the SEED database the performance measures are shown as the average over three sessions. A similar approach is also reported in previous studies [56], [57]. The improvement for EML algorithms over the CML algorithms shows that bagging followed by XGBoost and rotation forest algorithm can provide better emotion prediction compared to other ML-based algorithms. Bagging is the most optimal ML-based algorithm reporting the highest accuracy of 85.41% and 85.3%, for arousal and valence dimensions, respectively.

TABLE IV
PE-METRICS (AROUSAL) OF MULTICLASS EMOTION RECOGNITION FOR BAGGING ALGORITHM USING K10 PROTOCOL OVER NINE EMOTIONS FOR DREAMER DATABASE

PE Metric	Amuzement	Anger	Disgust	Excitement	Fear	Happy	Relax	Sad	Surprise	Overall PE
Sensitivity (%)	84.42	56.16	90.04	88.33	90.3	54.37	87.01	97.18	82.48	81.14
Specificity (%)	98.52	94.5	97.96	98.2	98.9	94.79	98.14	99.68	98.11	97.64
Kappa-score	0.84	0.51	0.86	0.86	0.9	0.5	0.84	0.97	0.81	0.79
F1-score (%)	86.04	56.1	87.26	87.17	90.7	55.42	86.18	97.31	83.42	81.07
AUC	0.98	0.93	0.99	0.99	0.99	0.93	0.98	1	0.98	0.97
ACC (%)	96.96	90.24	97.08	97.1	97.94	90.3	96.9	99.4	96.37	95.81

TABLE V
PE-METRICS (VALENCE) OF MULTICLASS EMOTION RECOGNITION FOR BAGGING ALGORITHM USING K10 PROTOCOL OVER NINE EMOTIONS FOR DREAMER DATABASE

PE Metric	Amuzement	Anger	Disgust	Excitement	Fear	Happy	Relax	Sad	Surprise	Overall PE
Sensitivity (%)	78	64.44	87.37	82.22	86.96	63.98	82.61	93.61	79.61	79.87
Specificity (%)	97.25	95.06	98.23	97.92	98.4	95.6	97.77	99.22	97.9	97.48
Kappa-score	0.75	0.58	0.85	0.81	0.85	0.6	0.8	0.93	0.79	0.77
F1-score (%)	77.99	63.19	86.69	82.69	87.06	64.24	82.43	93.69	81.06	79.89
AUC	0.97	0.95	0.99	0.98	0.99	0.95	0.98	0.99	0.98	0.98
ACC (%)	95.11	91.66	97.02	96.18	97.13	92.09	96.09	98.6	95.87	95.53

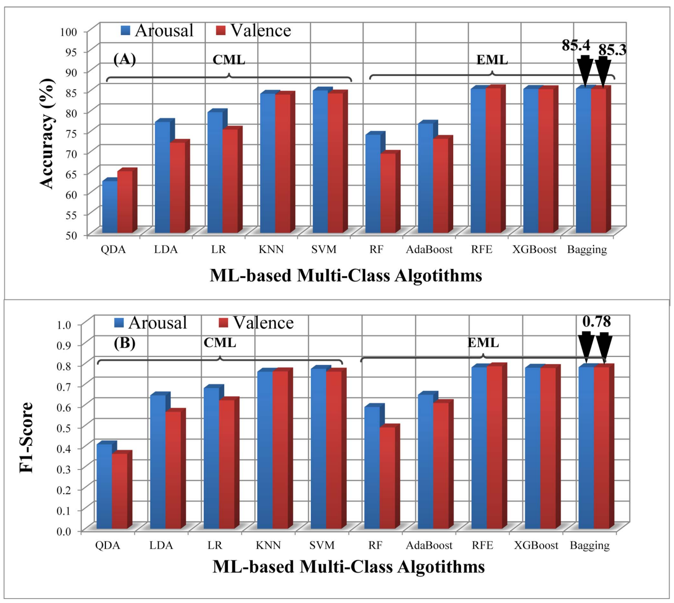


Fig. 7. Comparing the multiclass ML algorithms using (A) accuracy and (B) F1-score on SEED database.

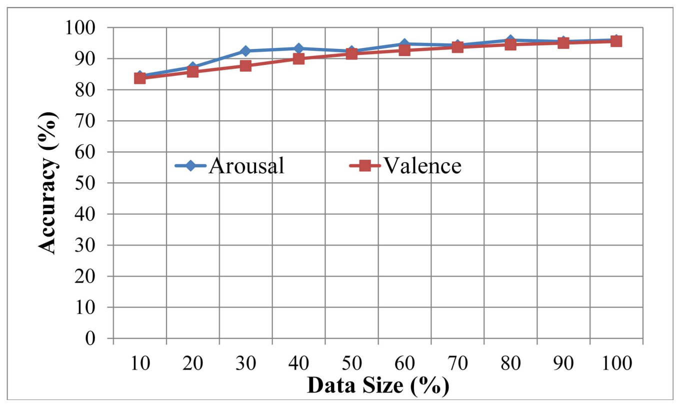


Fig. 8. Effect of data size on the performance of ML-based multiclass emotion recognition system.

The same is also true for other performance evaluation metrics of the bagging algorithm.

On the INTERFACES database, the mean accuracies over five EML algorithms are comparable with the mean accuracies over five CML algorithms for arousal (59.67% *vs.* 59.22%) and valence (59.67% *vs.* 59.22%) dimensions, respectively. However, the average F1-scores over the five EML algorithms are 99.80% (0.33 *vs.* 0.16) and 89.30% (0.38 *vs.* 0.20) higher compared to the average F1-scores over the five CML algorithms for both arousal and valence dimensions, respectively. On the MUSEC database, the mean accuracy over five EML algorithms are 9.90% (81.96% *vs.* 72.06%) and 3.58% (82.27% *vs.* 78.69%) higher compared to the mean accuracy over five CML algorithms for melody and song, respectively. Similarly, the mean F1-score over five EML algorithms are 104.68% (0.38 *vs.* 0.19) and 11.07% (0.70 vs. 0.63) higher compared to the mean F1-score over five CML algorithms for melody and song, respectively. The performance evaluation metrics for both INTERFACES and MUSEC databases are provided in Table XII and Table XIII from section IV of the supplementary material. Both the INTERFACES and MUSEC databases were highly imbalanced in terms of class distribution. Therefore, the F1-score can be considered as a reliable metric to compare the performance among the EML and CML algorithms. The higher performance metric in all the three validation databases validates our hypothesis that EML algorithms can provide better emotion recognition compared to the CML algorithms.

Fig. 8 investigates the effect of data size on the overall performance of a multiclass emotion recognition system. The accuracy for the proposed model slowly increases from 84.35% at 10% of the data size to 95.99% at 100% of the data size for arousal. Similarly, for the valence dimension, the

TABLE VI
COMPARING THE CML AND EML ALGORITHMS FOR MULTICLASS EMOTION RECOGNITION ON SEED DATABASE

Sr#	Type of Algorithm	Algorithm	Sensitivity (%)		Specificity (%)		Kappa-Score		F1-Score		AUC		Accuracy (%)	
			Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence	Arousal	Valence
1	CML	QDA	46.01	44.43	72.99	72.19	0.2	0.15	0.41	0.36	0.68	0.7	62.69	65.11
2		LDA	62.02	59.06	81.48	79.86	0.47	0.36	0.64	0.57	0.83	0.77	77.22	72.11
3		LR	66.51	64.31	83.54	82.22	0.53	0.44	0.68	0.62	0.85	0.83	79.62	75.35
4		KNN	75.56	76.17	87.64	87.83	0.64	0.64	0.76	0.76	0.92	0.91	84.15	83.93
5		SVM	77.44	76.71	88.52	88.17	0.67	0.64	0.77	0.76	0.92	0.91	84.95	84.21
		Average	**65.51**	**64.14**	**82.83**	**82.05**	**0.50**	**0.45**	**0.65**	**0.61**	**0.84**	**0.82**	**77.73**	**76.14**
6	EML	RF	58.27	53.86	79.38	76.85	0.39	0.28	0.59	0.49	0.79	0.75	74.07	69.42
7		AdaBoost	63.45	61.69	81.62	80.55	0.47	0.4	0.65	0.61	0.77	0.79	76.82	73.07
8		RFE	77.75	78.6	88.66	89.07	0.67	0.68	0.78	0.79	0.78	0.75	85.33	85.5
9		XGBoost	77.34	77.91	88.46	88.73	0.67	0.66	0.78	0.78	0.94	0.93	85.33	85.27
10		Bagging	77.29	78.07	88.47	88.8	0.66	0.67	0.78	0.78	0.94	0.93	85.41	85.3
		Average	**70.82**	**70.03**	**85.32**	**84.80**	**0.57**	**0.54**	**0.72**	**0.69**	**0.84**	**0.83**	**81.39**	**79.71**
Improvement for EML over CML(%)			**5.31**	**5.89**	**2.48**	**2.75**	**13.94**	**20.63**	**9.82**	**12.38**	**0.48**	**0.73**	**3.67**	**3.57**

TABLE VII
BENCH MARKING TABLE COMPARING THE PROPOSED STUDY WITH PREVIOUSLY PUBLISHED STUDIES ON THE DREAMER DATASET

SN	Author (Year)	Subjects	Feature Type	Decomposition Method	Algorithms	CV	Performance Evaluation Arousal	Valence
1	Katsigiannis et al. [37], (2018)	23	PSD	Bandpass Hamming sinc linear phase FIR filter	SVM	K10	Acc: 62.17%	Acc: 62.49%
2	Liu et al. [59], (2020)	23	Multi level features	Band pass filter	MLF-CapsNet, ConvReLU	K10	Acc: 95.26%	Acc: 94.59%
3	Cui et al. [60], (2020)	23	Temporal feature, DE, asymmetric features	Band pass filter	RACNN	K10	Acc:97.01±2.74%	Acc: 95.55±2.18%
4	Tao et al. [8], (2020)	23	Spatial and temporal features	Band pass filter	ACRNN	K10	Acc: 97.98±1.92%	Acc: 97.93±1.73%
5	Bhattacharyya et al. [58], (2021)	23	Spectral and temporal entropy	MFBSE-EWT	MFBSE-EWT, ARF	K10	Acc: 85.4%	Acc: 86.2%
6	Cheng et al. [55], (2021)	23	Spatio-temporal features	gcFOREST	Deep Forest	K10	Acc: 90.41±5.33%	Acc: 89.03±5.56%
7	Proposed method	23	Statistical features	DWT,EMD	Five CML, Five EML	K10	Acc: 93.79%	Acc: 94.5%

SN: serial number, PSD: power spectral density, DE: differential entropy, MFBSE-EWT: multivariate 4ier-Bessel series expansion empirical wavelet transform, gcFOREST: grained cascade forest, DWT: discrete wavelet transform, EMD: empirical mode decomposition, SVM: support vector machine, MLF-CapsNet: multi-level features guided capsule network, ConvReLU: convolutional rectified linear unit, RACNN: regional-asymmetric convolutional neural network, ACRNN: attention-based convolutional recurrent neural network, ARF: autoencoder random forest, FLDNet: frame-Level Distilling Neural Network, CML: conventional machine learning algorithm (SVM, linear discriminant analysis, logistic regression, quadratic discriminant analysis, K-nearest neighbors), EML: ensemble machine learning algorithm (random forest, rotation forest ensemble, AdaBoost, XGBoost, bagging), CV: cross-validation, Acc: accuracy

accuracy for the proposed model increases from 83.61% at 10% of the data size to 95.57% at 100% of the data size. Both the curves are almost similar to each other and indicate the need for more sample size to achieve the stabilize accuracy for the proposed model.

V. DISCUSSION

A. Claims and Summary

This study designed an ML-based system to predict multiclass emotions from the EEG signal. This is the first study that compares the performance of 10 multiclass ML algorithms for emotion recognition using EEG signals of the DREAMER database. Two categories of ML-based algorithms compared in this study are: (i) five CML algorithms (SVM, LR, LDA, QDA, and KNN) and (ii) five EML algorithms (RF, RFE, AdaBoost, Bagging, and XGBoost). The study results show that the ensemble algorithm such as bagging performed better in multiclass emotion recognition compared to the five types of CML algorithms. The classification accuracies for the bagging algorithm are 7.59% and 6.51% higher compared to the best conventional algorithm (i.e., KNN) for arousal and valence, respectively. The rotation forest and XGBoost also reported the higher performance in terms of accuracy, kappa score, and F1-score compared with CML algorithms. This has validated our hypothesis that EML-based algorithms can provide better multiclass emotion recognition compared to CML algorithms.

B. Benchmarking With Previous Studies

Table VII compares the proposed ML-based multiclass emotion recognition system with six previously published studies that use the DREAMER dataset. Most of the studies have used deep learning (DL)-based models to predict the emotions from the EEG signals. Similar to the proposed study, Bhattacharyya *et al.* [58] also used the concept of dual-decomposition using the MFBSE-EWT methods. However, their overall accuracy is 85.4% and 86.2% for arousal and valence, respectively. In Table VII, Katsigiannis and Ramzan [37] used the SVM-based ML system to predict the multiclass emotions and reported an overall accuracy of ~62%. The proposed method reported overall accuracy of 95.81% and 95.30% for arousal and valence dimensions. The performance of the proposed system was evaluated using the K10 cross-validation protocol, which is also popular and used in nearly all previous studies. This can be considered as the first ML-based study that used the bagging algorithm and compared its performance with nine other ML-based algorithms. The study results are encouraging since they are either better or comparable with the previously published ML or DL-based studies (Table VII).

C. Limitations and Future Extensions

Although the results obtained from the proposed model are better compared to previous studies on the same database, we still feel that a subject size of 23 and 14 channels are small. The study results need to validate on the larger database with more channels for EEG acquisition. This study used 26 statistical features for training and evaluating the proposed emotion recognition system. In the future, we intend to extend this study's results for investing the effect of time domain, frequency domain, and time-frequency domain features on the performance of the emotion recognition system. Another extension to this study could be the use of deep learning-based models such as deep neural networks and long short term memory networks that have shown promising results on previous EEG databases.

VI. Conclusion

This study compared the performance of five ensemble-based ML algorithms with five conventional ML algorithms for recognizing nine emotions from EEG signals. A dual decomposition technique, which is combination of DWT and EMD methods, is adopted for decomposition of EEG signals methods. In the present study, it has been found that all five ensemble-based ML algorithms provide better emotion recognition compared to conventional ML algorithms, supporting the fact that the ensemble learning-based algorithms can be used over conventional ML algorithms for EEG-driven emotion recognition. Out of five ensemble algorithms, bagging algorithm is found to be superior over other ML-based algorithms with 95.81% and 95.53% as the emotion recognition accuracy for both arousal and valence dimensions, respectively. The proposed ML-based system is verified using three external databases that also confirmed our hypothesis that the ensemble learning-based algorithms are better in emotion recognition compared to the conventional ML-based algorithms. This study used the ANOVA-based feature selection, which has shown an improvement in the overall emotion recognition ability of the proposed ensemble learning-based ML system.

Acknowledgment

The authors of this manuscript are thankful to Prof. Theerawit Wilaiprasitporn and his team from the Bio-Inspired Robotics and Neural Engineering Laboratory, School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand, for providing the MUSEC and INTERFACES databases. The authors are also thankful to Prof. Bao-Liang Lu from the Center for Brain-like Computing and Machine Intelligence for making available the SEED database.

References

[1] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017.

[2] Y. Yang, Q. M. J. Wu, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2017.

[3] Z. Li, S. Zhao, J. Duan, C.-Y. Su, C. Yang, and X. Zhao, "Human cooperative wheelchair with brain–machine interaction based on shared control strategy," *IEEE/ASME Trans. Mechtronics*, vol. 22, no. 1, pp. 185–195, Feb. 2017.

[4] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.

[5] G. F. Wilson and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 45, no. 4, pp. 635–644, Dec. 2003.

[6] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychol.*, vol. 14, no. 4, pp. 261–292, 1996.

[7] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Trans. Affect. Comput.*, early access, May 14, 2019, doi: 10.1109/TAFFC.2019.2916015.

[8] W. Tao *et al.*, "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, early access, Sep. 22, 2020, doi: 10.1109/TAFFC.2020.3025777.

[9] D. Maheshwari, S. K. Ghosh, R. K. Tripathy, M. Sharma, and U. R. Acharya, "Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel EEG signals," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104428.

[10] S. Sangnark *et al.*, "Revealing preference in popular music through familiarity and brain response," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14931–14940, Jul. 2021.

[11] V. Bajaj, S. Taran, and A. Sengur, "Emotion classification using flexible analytic wavelet transform for electroencephalogram signals," *Health Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–7, Dec. 2018.

[12] S. K. Khare and V. Bajaj, "Constrained based tunable Q wavelet transform for efficient decomposition of EEG signals," *Appl. Acoust.*, vol. 163, Jun. 2020, Art. no. 107234.

[13] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[14] S. Taran, V. Bajaj, D. Sharma, S. Siuly, and A. Sengur, "Features based on analytic IMF for classifying motor imagery EEG signals in BCI applications," *Measurement*, vol. 116, pp. 68–76, Feb. 2018.

[15] R. Patel, M. P. R. Janawadkar, S. Sengottuvel, K. Gireesan, and T. S. Radhakrishnan, "Suppression of eye-blink associated artifact using single channel EEG data by combining cross-correlation with empirical mode decomposition," *IEEE Sensors J.*, vol. 16, no. 18, pp. 6947–6954, Sep. 2016.

[16] A. Bhattacharyya, M. Sharma, R. B. Pachori, P. Sircar, and U. R. Acharya, "A novel approach for automated detection of focal EEG signals using empirical wavelet transform," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 47–57, 2018.

[17] T. Tuncer, S. Dogan, and A. Subasi, "A new fractal pattern feature generation function based emotion recognition method using EEG," *Chaos, Solitons Fractals*, vol. 144, Mar. 2021, Art. no. 110671.

[18] P. Lakhan *et al.*, "Consumer grade brain sensing for emotion recognition," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9896–9907, Nov. 2019.

[19] V. Gupta, M. D. B. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2266–2274, Mar. 2019.

[20] L. D. Sharma and A. Bhattacharyya, "A computerized approach for automatic human emotion recognition using sliding mode singular spectrum analysis," *IEEE Sensors J.*, vol. 21, no. 23, pp. 26931–26940, Dec. 2021.

[21] S. K. Khare and V. Bajaj, "An evolutionary optimized variational mode decomposition for emotion recognition," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2035–2042, Jan. 2021.

[22] 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. 29th IAAI Conf.*, 2017, pp. 4746–4752.

[23] H. Chao and L. Dong, "Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2024–2034, Jan. 2021.

[24] F. Demir, N. Sobahi, S. Siuly, and A. Sengur, "Exploring deep learning features for automatic classification of human emotion using EEG rhythms," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14923–14930, Jul. 2021.

[25] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1, Aug. 1995, pp. 278–282.

[26] T. Chen and T. He, "XGBoost: EXtreme gradient boosting," *R Package Version 0.4-2*, vol. 1, no. 4, pp. 1–4, Aug. 2015.

[27] C. Wei, L.-L. Chen, Z.-Z. Song, X.-G. Lou, and D.-D. Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101756.
[28] A. Gicić and A. Subasi, "Credit scoring for a microcredit data set using the synthetic minority oversampling technique and ensemble classifiers," *Expert Syst.*, vol. 36, no. 2, Apr. 2019, Art. no. e12363.
[29] A. Subasi *et al.*, "Sensor based human activity recognition using AdaBoost ensemble classifier," *Proc. Comput. Sci.*, vol. 140, pp. 104–111, Jan. 2018.
[30] A. Subasi, T. Tuncer, S. Dogan, D. Tanko, and U. Sakoglu, "EEG-based emotion recognition using tunable *Q* wavelet transform and rotation forest ensemble classifier," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102648.
[31] A. B. Das and M. I. H. Bhuiyan, "Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain," *Biomed. Signal Process. Control*, vol. 29, pp. 11–21, Aug. 2016.
[32] R. Sharma, R. Pachori, and U. Acharya, "An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures," *Entropy*, vol. 17, no. 12, pp. 5218–5240, Jul. 2015.
[33] L. Yuan, B. Yang, S. Ma, and B. Cen, "Combination of wavelet packet transform and Hilbert–Huang transform for recognition of continuous EEG in BCIs," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, Aug. 2009, pp. 594–599.
[34] N. Ji, L. Ma, H. Dong, and X. Zhang, "EEG signals feature extraction based on DWT and EMD combined with approximate entropy," *Brain Sci.*, vol. 9, no. 8, p. 201, Aug. 2019.
[35] S. Taran and V. Bajaj, "Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method," *Comput. Methods Programs Biomed.*, vol. 173, pp. 157–165, May 2019.
[36] V. Gupta, A. Bhattacharyya, and R. B. Pachori, "Classification of seizure and non-seizure EEG signals based on EMD-TQWT method," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, Aug. 2017, pp. 1–5.
[37] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.
[38] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Aug. 27, 2019, doi: 10.1109/TAFFC.2019.2937768.
[39] A. L. Tandle, M. S. Joshi, A. S. Dharmadhikari, and S. V. Jaiswal, "Mental state and emotion detection from musically stimulated EEG," *Brain Informat.*, vol. 5, no. 2, pp. 1–13, Dec. 2018.
[40] J. D. Kropotov, *Quantitative EEG, Event-Related Potentials and Neurotherapy*. New York, NY, USA: Academic, 2010.
[41] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.
[42] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.
[43] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals," *Digit. Signal Process.*, vol. 78, pp. 185–196, Jul. 2018.
[44] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[45] V. Doma and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *J. Big Data*, vol. 7, no. 1, pp. 1–21, Dec. 2020.
[46] M. R. Hasan, M. I. Ibrahimy, S. Motakabber, and S. Shahid, "Classification of multichannel EEG signal by linear discriminant analysis," in *Progress in Systems Engineering*. Cham, Switzerland: Springer, 2015, pp. 279–282.
[47] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. N. Tibarewala, "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data," in *Proc. Int. Conf. Syst. Med. Biol.*, Dec. 2010, pp. 126–131.
[48] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
[49] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[50] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class AdaBoost," *Statist. Interface*, vol. 2, no. 3, pp. 349–360, 2009.
[51] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Jul. 2010.
[52] J. Brownlee. *Nested Cross-Validation for Machine Learning With Python*. Accessed: Nov. 1, 2020. [Online]. Available: https://machinelearningmastery.com/nested-cross-validationfor-machine-learning-with-python/
[53] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[54] F. Pedregosa, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.
[55] J. Cheng *et al.*, "Emotion recognition from multi-channel EEG via deep forest," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 453–464, Feb. 2021.
[56] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.
[57] G. Zhang, M. Yu, Y.-J. Liu, G. Zhao, D. Zhang, and W. Zheng, "SparseDGCNN: Recognizing emotion from multichannel EEG signals," *IEEE Trans. Affect. Comput.*, early access, Jan. 13, 2021, doi: 10.1109/TAFFC.2021.3051332.
[58] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, "A novel multivariate-multiscale approach for computing EEG spectral and temporal complexity for human emotion recognition," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3579–3591, Feb. 2021.
[59] Y. Liu *et al.*, "Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103927.
[60] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106243.

Kranti S. Kamble (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and telecommunication engineering from the Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra, India, in 2012, and the M.Tech. degree in electronics and telecommunication engineering from the Veermata Jijabai Technological Institute, Mumbai, India, in 2015. She is currently pursuing the Ph.D. degree with the Visvesvaraya National Institute of Technology Nagpur, Nagpur, India. Her research interests include biomedical signal processing and artificial intelligence.

Joydeep Sengupta received the B.Tech. degree from the University of Kalyani, Kalyani, India, in 2002, the M.Tech. degree from The University of Burdwan, Bardhaman, India, in 2004, and the Ph.D. degree from the Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India (formerly Bengal Engineering and Science University, Shibpur), in 2017. He is currently an Assistant Professor with the Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology Nagpur, Nagpur, India. He has authored or coauthored several research papers in reputed international journals and conferences. His research interests include microwave engineering, antenna design, and optical communication.