Enhancing EEG-based emotion recognition using Semi-supervised Co-training Ensemble Learning

Rachel Yeo Hui Min
School of Computing,
National University of Singapore
Singapore, Singapore
rachelyeohm@u.nus.edu

Aung Aung Phyo Wai
Center for Brain Computing Research,
School of Computer Science and Engineering
Nanyang Technological University
Singapore, Singapore
apwaung@ntu.edu.sg

***Abstract*— Emotion recognition based on Brain-Computer Interface is crucial in deepening our understanding of humans' emotions and decision-making process. The enhanced precision in emotion measurement allows for more rigorous analysis of mental disorders and therapy effectiveness. Our method aims to solve two challenges in this field. First, existing models' features often fail to comprehensively capture the multiple dimensions of information in EEG signals, like temporal or frequency domains. We propose two models: a CNN model trained on temporal features and a DNN model trained on differential entropy features, and ensemble their predictions with weighted voting. Second, labels can be uncertain, where data is unconfidently labelled. This is due to emotions' subjectivity causing a lack of clear ground truth in EEG. The proposed method aims to mitigate this by using a semisupervised method that utilises data with uncertain labels as unlabelled data. Co-training is used to allow the two models to learn from each other. Our combined model achieves higher accuracy than temporal and spectral models by 8.51% and 6.87% respectively for the SEED dataset. For the MER dataset, its accuracy outperforms temporal and spectral models by 24.96% and 2.70% respectively for arousal classification, and 11.18% and 49.21% respectively for valence classification.**

Keywords*— Brain-Computer Interface, Emotion recognition, Semi-supervised learning, Ensemble learning, Multiview learning, Deep learning

I. Introduction

Emotion recognition using a Brain-Computer Interface is a scientific way of measuring emotions, and it allows for precise, less subjective ways to study mental disorders and evaluate the effectiveness of treatment methods. Through this, valuable insights can also be gained about how emotions work in the brain.

Emotions are captured using EEG signals. However, the complexity of EEG signals means that even medical experts face challenges in manually crafting features to decipher emotions from EEG signals [1]. Thus, artificial intelligence and deep learning have emerged as a natural solution, being able to identify patterns within EEG signals.

One challenge in EEG is fully capturing the various dimensions of information found in EEG signals, like temporal, spectral and spatial domains. Yet, most feature extraction methods extract features corresponding to only one domain, causing an incomplete representation of EEG data. To mitigate this, two models are trained, one on temporal features and one on spectral features. Both types of features have been shown to be effective [2, 3, 4] in emotion recognition models. This allows each model to focus on extracting patterns from its specific feature space. Such an approach can result in higher accuracy and lower classification errors than inputting all features into a single classifier. [5, 6] Ensemble learning, in particular weighted voting, is used to combine the predictions of the two classifiers to generate more accurate final predictions [7, 8]

Emotion recognition faces another challenge of uncertain labels, where data is not confidently or accurately labelled. This is exacerbated by the subjective nature of emotion. The absence of a clear ground truth makes it challenging to determine whether emotion labels are accurate [9, 10].

This subjectivity happens in multiple ways throughout the EEG experimental process. In the design phase, stimuli used to elicit emotions may not evoke the correct emotional feedback from every subject, as different subjects react differently to the same stimuli. Subjects becoming mentally fatigued throughout the experiment could cause the recorded emotions to differ from the emotion label [10]. Intra-subject and inter-subject variability in identifying emotions can also make self-reported emotion labels uncertain. These factors all decrease classification performance. [11, 12, 13]

Various solutions have been proposed to deal with the issue of uncertain labels in EEG, but they largely do not address the problem sufficiently. Robust classifiers are less affected by noise [14, 15], but prioritising robustness to label noise has an inherent trade-off with accuracy and can worsen performance [11]. Completely removing data with uncertain labels would discard valuable, limited data [16]. There is thus a research gap where there are few methods that alleviate EEG label uncertainty effectively while maximising the usage of data.

Our solution addresses this gap. The labels of the noisiest data are removed, and a semi-supervised model is trained, with the uncertain data being reused as unlabelled data. This allows uncertain data to still be used without corrupting the models with potential label noise. After co-training, the predictions of the two classifiers are combined using weighted voting to generate the final predictions.

Our major contributions are listed as such:

1. We develop a novel hybrid co-training and ensemble method that effectively mitigates the issue of uncertain labels in EEG.
2. We combine different representations of EEG signals by allowing models trained on different features to learn from each other. This forms a more complete understanding of the data.
3. Our method of combining different models together consistently enhances performance compared to the individual models, and can be used on top of base models to provide relative improvements in accuracy.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00099

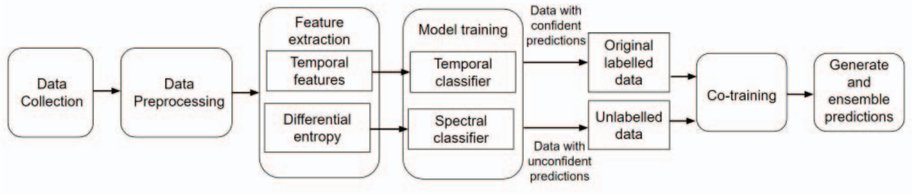

Fig. 1. Summary of Methodology.

4. More generally, our emotion recognition model can be used in healthcare to assess patients with mental health conditions and can be made part of assistive communication devices.

II. Related Work

A. Temporal features

Statistical features are a reliable method of EEG emotion classification. Liu and Sourina [4] extracted 6 common statistical features, including Higher Order Crossing features, from the DEAP dataset. Using a SVM classifier, they obtained a mean accuracy of 85.38%. Rahman et al. [3] extracted 10 statistical features, including L2-norm and fractal dimension. When using Artifical Neural Network (ANN), they achieved 86.57% on the 3-class SEED dataset.

B. Non-linear features

Non-linear features are another representation of EEG signals that have been effective in emotion classification. This includes complexity features like differential entropy and Higuchi Fractal Dimension. Iyer et.al. [2] extracted differential entropy features and used the ensemble of three models, CNN, LSTM and CNN-LSTM to classify emotions. They attained 97.16% accuracy on the SEED dataset.

C. Spectral features

EEG signals also consist of a frequency domain that can be taken advantage of to generate useful features. For example, it is common to decompose EEG signals to smaller frequency bands and calculate features like Power Spectral Density, average bandpower and differential entropy.

Rahman et. al. [17] generated topographic images from RPSD features: the ratio of the Power Spectral Density (PSD) of the band of interest to the PSD of the total frequency band. Using a CNN, they attained an average classification accuracy of 94.63% among all the subjects.

D. Multi-view learning and semi-supervised learning

Multi-view learning and semi-supervised learning have recently been used in EEG emotion classification to improve performance over single-view learning and fully-supervised learning respectively.

Gao, Fu, Ouyang and Wang [18] used a spatio-temporal and self-adaptive GCN to combine spatial and temporal domain information into one model. Additionally, differential asymmetry (DASM) and rational asymmetry (RASM) spectral features were also used in the model. They attained an accuracy of 86.00% on the SEED dataset.

Zhang, Davoodina and Etemad [19] proposed PARSE, a semi-supervised model which could learn on large amounts of unlabelled data and limited amounts of labeled data. This was done through data augmentation, label guessing of unlabelled data, and refining the guessed labels. They attained 91.14% accuracy when there were 25 instances of labelled data.

III. Methodology

Fig. 1 summarises the methodology process.

A. Datasets

The SJTU SEED dataset [20, 21] is an emotion dataset which comprises 62-channel EEG signals and eye movement data, collected using the ESI NeuroScan System and SMI eye-tracking glasses respectively. It was generated by showing 15 film clips as stimuli. There were 15 subjects, and each subject repeated the trial 3 times. The film clips were around 4 minutes each, and are labelled positive, neutral and negative based on their content.

The Music Emotion Regression (MER) dataset is a dataset constructed by recording the EEG signals of 50 participants exposed to music stimuli. Each participant goes through two experiment sessions. In each session, the subject is briefed, goes through a practice run, and then undergoes 3 blocks of experiments with rests in between. Each block consists of 13 trials - in each trial, a subject listens to a music clip, chooses their arousal and valence rating, and rests for 15 seconds. The music used are generated by a deep learning model, constructed to elicit specific valence and arousal levels. There are two emotion recognition tasks for this dataset, identifying high / low arousal levels and high / low valence levels.

B. Preprocessing

For the SEED dataset, the collected signals were downsampled to 200Hz, and a bandpass filter of 0 - 75Hz was applied to filter out low-frequency drifts and high-frequency noise.

For the MER dataset, bandpass filtering from 0.3 to 4Hz and downsampling from 1000Hz to 250Hz was conducted on the dataset. Afterwards, the Common Average Referencing (CAR) rereferencing technique was used, where the average of the signal of all electrodes was subtracted from the EEG signal, to filter out common noise. Independent Component Analysis and EOG artifact removal was then conducted to remove noisy signal components like those generated from blinking.

All signals in the dataset are cropped to match the length of the shortest signal. A sliding window is applied on the signal to generate 4 seconds windows with a window overlap of 2 seconds. This maximises the available data.

1) Extracting temporal features: For each window, temporal features of skew, kurtosis, Higuchi Fractal Dimension (HFD), and hjorth parameters like hjorth activity, mobility and complexity were extracted.

Fractal Dimension is a geometric method of representing complexity in the temporal domain. Being non-linear, it works for non-stationary signals like EEG signals, unlike methods like Fourier Transform which require the assumption of stationarity. HFD, a variation of the box-counting fractal dimension, is used is used to approximate Fractal Dimension, as it provide closer approximations of theoretical values than other methods *[22, 23]*.

Let a discrete time series be represented as X : $x(1), x(2), \ldots, x(N)$ where N is the total number of data samples. From this series, k new time series are constructed as such.

$$X_m(k) = x(m), x(m+k), x(m+2k), \ldots, x\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor\right) \quad (1)$$

$$m \in \{1, \ldots, k\} \quad (2)$$

$$k \in \{1, \ldots, k_{max}\} \quad (3)$$

$$k_{max} \geq 2 \quad (4)$$

where m is the initial time, k is the time interval, and $\lfloor\ \rfloor$ is the floor function, and X_m is the discrete time series for initial time m .

From here, the length of each curve, represented as $L_m(k)$, is calculated.

$$L_m(k) = \frac{1}{k} \sum_{i=1}^{\left\lfloor \frac{N-m}{k} \right\rfloor} |x(m+ik) - x(m+(i-1)k)| \frac{N-1}{\left\lfloor \frac{N-m}{k} \right\rfloor k} \quad (5)$$

Then, the length of the time interval k, represented as L_k, is as follows:

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_m(k) \quad (6)$$

The HFD is the slope of the best-fit linear plot of the data points $log\,\frac{1}{k}$ and $\log L(k)$.

Hjorth parameters are also utilised, which have been used in EEG emotion recognition to good effect [24, 25].

For a signal of $x(t)$ of N datapoints, the formulas for Hjorth activity, mobility and complexity are as shown:

$$Activity_x = var(x(t)) = \frac{1}{N} \sum_{n=1}^{N} |x(t) - \mu_x|^2 \quad (7)$$

$$Mobility_x = \sqrt{\frac{var(\frac{dx(t)}{dt})}{var(x(t))}} \quad (8)$$

$$Complexity_x = \frac{Mobility(\frac{dx(t)}{dt})}{Mobility(x(t))} \quad (9)$$

Trials were conducted to determine the effectiveness of these features in improving classification accuracy, and features with a positive impact on the accuracy were used in the final model.

2) Extracting spectral complexity features: For our second view, EEG signals are decomposed into 4 sub frequency bands, and non-linear differential entropy features are extracted.

Firstly, a bandpass filter was used to decompose and filter the signal to produce theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz) and gamma (30-64Hz) signals. This is important in the process of extracting differential entropy features, as for a fixed length EEG sequence, an EEG signal will follow a Gaussian distribution once it is decomposed to smaller frequency bands *[26]*. Thus, the signal X of a certain frequency band then conforms to normal distribution $X \sim N(\mu, \sigma^2)$, and the probability density function of the signal can be expressed as a function of the mean μ and the standard deviation σ. This will simplify subsequent analysis.

Differential entropy is based on the concept of Shannon entropy from information theory. For a continuous random variable, Shannon entropy is defined as:

$$Entropy = -\int_{-\infty}^{+\infty} p(x) \log(p(x))\, dx \quad (10)$$

where $p(x)$ represents the probability density function of the signal. The probability density function can then be simplified as shown below:

$$DE = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) \quad (11)$$

$$= \frac{1}{2} \log(2\pi e \sigma^2)$$

where DE is the differential entropy of the signal.

Differential entropy features are then combined with channel data to obtain a feature vector of size 248 for each sample. The feature data is scaled to a range of 0 to 1. Finally, Principal Component Analysis (PCA) is applied to shrink the feature vector down to 100 components. This minimises redundancies in the extracted features.

C. Model Architecture

1) Temporal view model: A CNN model architecture is used to model the selected temporal features. However, 2D-CNN models require data samples with 2 spatial dimensions. For our data with 1 spatial dimension, a 1D-CNN variant was found to be more suitable, and is also capable of extracting valuable temporal information from data *[27]*. Accordingly, 1D versions of convolutional layers and max pooling layers are used. The model architecture consisted of 4 sets of a convolutional layer followed by max-pooling layer. After every two convolutional layers, a dropout layer was added. A dropout value of 0.1 was empirically chosen. Additionally, every convolutional layer was followed by a Rectified Linear Unit (ReLU) activation function to strongly confine the outputs of the weights to positive values. Two dense layers and a softmax activation function are applied at the end of the model architecture. The model was trained for 60 with a batch size of 32.

2) Spectral view model: To train the differential entropy features, a Deep Dense Neural Network is used. Dropout layers are added to help address the issue of overfitting, with p = 0.2. The DNN consists of a set of 2 dense layers followed by 1 dropout layer, then another set of the same layers. An Adam optimizer and categorical cross entropy loss was used in the training of the model. The

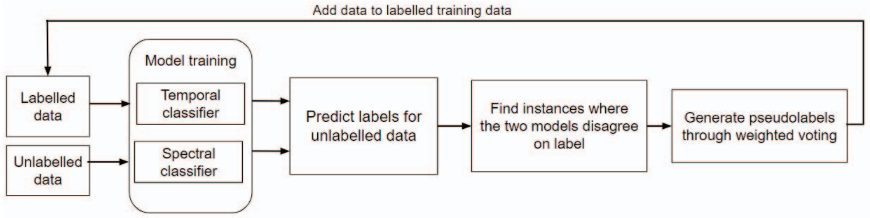

Fig. 2. Diagram of co-training process

Model	Accuracy	Precision	Recall	F1
Temporal Model [28]	0.9004	0.9018	0.9004	0.9006
Spectral Complexity Model [29]	0.9142	0.9141	0.9142	0.9140
Ensembled model made from temporal and spectral model [30]	0.9528	0.9528	0.9527	0.9528
Single classifier trained on temporal and spectral features [31]	0.9429	0.9437	0.9429	0.9430
Co-training Ensembled Temporal and Spectral Model	0.9770	0.9771	0.9769	0.9769

TABLE I. RESULTS OF SEED DATASET

model was trained for 160 epochs until convergence, with a batch size of 32.

3) Co-Training: Our two base models will then be used with a co-training paradigm. For the proposed use case, the original co-training paradigm has been adapted to focus on disagreement between the two base models.

The process of co-training is as follows: In each iteration of co-training, both base models are trained until accuracy stabilises. Then, the models are used to predict class probabilities for unlabelled data instances. A subset of data where the two models predict different labels is then added into the training dataset with replacement. Pseudolabels are generated by ensembling the predictions of the two models together to create a stronger prediction. Each round, the training dataset will consist of the original labelled dataset and a subset of data that the models disagreed on the previous round. This cycle occurs until the combined model converges. Fig. 2 summarises the process of co-training.

This approach is slightly different from the original co-training method, which adds data with high confidence predictions to the training dataset [32]. However, adding examples which models disagree on to the training dataset allows the models to learn from each other. Additionally, instances that models disagree on are likely to be instances of uncertain labels.

D. Training Methodology

To assess both the performance and reliability of our proposed method, k-fold cross-validation is used, with k = 5. This involves separating the dataset into 5 subsets, training the model on 4 (k-1) subsets, and validating it on the remaining subset. This process is repeated 5 times with a different subset being used for validation each time. With this, the model's ability to generalise to new unseen data and the variability of its performance can be measured.

IV. RESULTS AND ANALYSIS

In this section, the performance of the individual temporal and spectral models and the overall co-training ensembled model are evaluated for both datasets. To analyse the benefits of ensembling, the performance of the individual CNN and DNN models, without ensembling or co-training, is also computed. The parameters of these models follow those listed in Section III-C-1 and Section III-C-2. The performance of the ensembled predictions of these two models without co-training is also computed and compared to that of our proposed method.

A. Results for SEED Dataset

Firstly, from Table I, the individual temporal model and spectral model achieve good results on their own. The model trained on temporal features achieves an accuracy of 90.04% and a F1 score of 90.06%, and the model trained on differential entropy features achieve an accuracy of 91.42% and a F1 score of 91.40%. This is indicative of the suitability of the model architecture and features selected to this emotion recognition task.

Still, when comparing individual models to the proposed co-training ensembled model, the proposed method outperforms both individual models significantly. Our method attained 97.70% accuracy and an F1 score of 97.69%. This is a 8.51% and 6.87% improvement in accuracy over the temporal and spectral models respectively.

Additionally, ensembling predictions of both individual models without using co-training also outperformed individual models, demonstrating the effectiveness of ensembling itself in improving performance. The accuracy of these ensembled predictions are higher than that of the temporal and spectral models, by 5.82% and 4.22% respectively. However, the co-training ensemble method still performed the best, showing the benefits of our proposed approach.

B. Results for MER Dataset

For the MER Dataset, emotion recognition is split into an arousal classification task and a valence classification task. The results for both tasks can be found in Table II.

For the arousal task, the spectral classifier performs significantly better than the temporal classifier, attaining 75.52% accuracy compared to 62.07% accuracy for the temporal model. But the proposed method still shows a performance increase, achieving an accuracy of 77.56%.

	Arousal				Valence			
Model	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**
Temporal Model [28]	0.6207	0.6229	0.6213	0.6192	0.8604	0.8627	0.8618	0.8618
Spectral Complexity Model [29]	0.7552	0.7563	0.7552	0.7549	0.6411	0.6439	0.6428	0.6421
Ensembled model made from temporal and spectral model [30]	0.7705	0.7705	0.7704	0.7704	0.8818	0.8827	0.8820	0.8820
Single classifier trained on temporal and spectral features [31]	0.6549	0.6560	0.6540	0.6534	0.9386	0.9387	0.9386	0.9386
Co-training Ensembled Temporal and Spectral Model	0.7756	0.7768	0.7766	0.7766	0.9566	0.9595	0.9594	0.9594

TABLE II. RESULTS OF MER DATASET, FOR AROUSAL (LEFT) AND VALENCE (RIGHT)

The valence task finds the temporal classifier performing much better than the spectral classifier, where the temporal model attains 86.04% accuracy and the spectral model achieves 64.11% accuracy. The proposed method also outperforms the base models significantly, achieving an accuracy of 95.66%.

Similar trends in performance are observed in the MER dataset as in the SEED dataset. Ensembling predictions of the base models together without co-training also attains superior performance than the individual models. For the arousal task, it shows a 24.13% and 2.03% increase in accuracy over the temporal and spectral model respectively. For the valence task, it demonstrates a 2.59% and 37.54% increase in accuracy compared to the temporal and spectral model respectively.

V. DISCUSSION

A. *The effectiveness of ensembling*

A major requirement for the effectiveness of ensemble learning is diversity in the feature selection and/or model architecture is strongly encouraged [32], such that combining both classifiers can leverage both classifiers' strengths to improve performance and compensate for individual models' weaknesses. In aiming to fulfil this requirement, the proposed method has ensured the two sets of features used to train the models have been chosen to be suitably distinct, being taken from different domains of the EEG signal. Different model architectures were also deliberately chosen for the two classifiers. This seems to be an effective method to ensure diversity, as our proposed method shows an improvement in performance.

However, one possible concern about ensembling different feature sets together is that a single classifier trained on all the feature sets could provide similar performance, rendering ensemble learning techniques superfluous in this case. To investigate this, a single classifier was also trained on both temporal and spectral features, and its performance will be compared the proposed method. For the SEED dataset, this single classifier attained an accuracy of 94.29% and a F1 score of 94.30%, lower than the proposed method. For the MER dataset's arousal and valence tasks, the single classifier also achieved a lower accuracy and F1 score than the proposed. This could be because the relatively high dimensionality of all the features together could require a model with increased complexity and more data to model accurately. While the performance of both methods could be different for different scenarios, ensemble learning's superior performance here demonstrates the benefits of ensembling in this case.

B. *The effectiveness of co-training*

Much research has been done to investigate the sufficient conditions for co-training to be successful. One such condition is that the feature sets are class conditionally independent [33]. However, it is rather unlikely that the temporal and spectral features used in the proposed method would be class conditionally independent. In fact, as both temporal and spectral features are generated from the same EEG signal, there is a level of interdependence between them. It has also been shown that weak dependence can also guarantee the success of co-training (bootstrapping), but this is rather challenging to show. Since the effectiveness of co-training in EEG cannot be discerned easily, the best way to determine it is to test it out experimentally, which is the motivation behind the proposed method. Nonetheless, even while the feature sets may not be class-conditional-independent or weakly dependent, the experimental results have shown that co-training is still effective in our case.

Co-training was also intended to mitigate the issue of uncertain labels. Since uncertain labels contaminate the dataset with noise and reduce accuracy, the improvement in performance when co-training is used, compared to simple ensembling, indicates that this method is successful.

VI. FUTURE WORK

The feasibility of the proposed method has been demonstrated for model architectures like CNN and DNN for the base models. However, future work could include testing the proposed method on different model architectures like LSTM, Attention Networks and Graph Neural Networks.

This framework utilised weighted voting specifically to demonstrate the effectiveness of ensembling in improving performance. However, more work could be done to test the performances of different ensembling methods, like bagging, boosting and stacking, in order to determine the optimal method for ensembling models in this scenario.

Finally, future work could include comparison and evaluation with other state-of-the-art methods.

VII. CONCLUSION

The proposed semi-supervised co-training ensembled model is effective in increasing model performance over individual models, as shown as its better overall performance compared to individual models for both SEED and MER datasets.

It also mitigates the issue of models not fully representing EEG data, by ensembling classifiers trained on different features from different domains of the EEG signals. This method consistently outperforms training a single classifier on all the features.

The findings also suggest that semisupervised co-training is effective in mitigating the issue of uncertain labels and improves performance.

Therefore, the proposed method shows promise in enhancing the performance of EEG emotion recognition and mitigating some of the common challenges faced by emotion recognition tasks.

VIII. References

[1] S. Gong, K. Xing, A. Cichocki and J. Li, "Deep learning in EEG: Advance of the last ten-year critical period," *IEEE Transactions on Cognitive and Developmental Systems,* vol. 14, no. 2, pp. 348-365, 2022.

[2] A. Iyer, S. S. Das, R. Teotia, S. Maheshwari and R. R. Sharma, "CNN and LSTM based Ensemble Learning for human emotion recognition using EEG Recordings," *Multimedia Tools and Applications,* vol. 82, no. 4, p. 4883–4896, Apr. 2022.

[3] A. Rahman, F. Hossain, M. Hossain and R. Ahmmed, "Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal," *Egyptian Informatics Journal,* vol. 21, no. 1, pp. 23-35, Mar. 2020.

[4] Y. Liu and O. Sourina, "EEG-based subject-dependent emotion recognition algorithm using fractal dimension," *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* vol. 3171, p. 3166, Oct. 2014.

[5] F. Lotte, M. Congedo, A. L.´ecuyer, F. Lamarche and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of Neural Engineering,* vol. 4, no. 2, 2007.

[6] A. Soria-Frisch, "A Critical Review on the Usage of Ensembles for BCI," *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications,* pp. 41-65, 2013.

[7] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 8, no. 4, 2018.

[8] L. Li, Q. Hu, X. Wu and D. Yu, "Exploration of classification confidence in ensemble learning," *Pattern recognition,* vol. 47, no. 9, p. 3120–3131, 2014.

[9] Y. Höller, J. Bergmann, A. Thomschewski, M. Kronbichler, P. H. H¨oller, J. S. Crone, E. V. Schmid, K. Butz, R. Nardone and E. Trinka, "Comparison of EEG-Features and Classification Methods for Motor Imagery in Patients with Disorders of Consciousness," *PLOS ONE,* vol. 8, no. 11, 2013.

[10] A. K. Porbadnigk, N. Gornitz, M. Kloft and K.-R. Muller, "Decoding Brain States during Auditory Perception by Supervising Unsupervised Learning," *Journal of Computing Science and Engineering,* vol. 7, no. 2, pp. 112-121, 2013.

[11] B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 25, no. 5, pp. 845-869, 2013.

[12] J. Li and L. Zhang, "Active training paradigm for motor imagery BCI," *Experimental Brain Research,* vol. 219, no. 2, pp. 245-254, Apr. 2012.

[13] M. Gao, R. Liu and J. Mao, "Noise robustness low-rank learning algorithm for electroencephalogram signal classification," *Frontiers in Neuroscience,* vol. 15, 2021.

[14] M. Pechenizkiy, A. Tsymbal, S. Puuronen and O. Pechenizkiy, "Class noise and supervised learning in medical domains: The effect of feature extraction," *19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06),* p. 708–713, 2006.

[15] P. Zhong, D. Wang and C. Miao, "EEG-Based Emotion Recognition Using Regularized Graph Neural Networks," *IEEE Transactions on Affective Computing,* vol. 13, no. 3, p. 1290–1301, 2020.

[16] L. Xu, M. Xu, Y. Ke, X. An, S. Liu and D. Ming, "Cross-Dataset Variability Problem in EEG Decoding With Deep Learning," *Frontiers in Human Neuroscience,* vol. 14, 2020.

[17] A. Rahman, A. Anjum, M. H. Milu, F. Khanam, M. S. Uddin and N. Mollah, "Emotion recognition from EEG-based relative power spectral topography using convolutional neural network," *Array,* vol. 11, Sept. 2021.

[18] Y. Gao, X. Fu, T. Ouyang and Y. Wang, "EEG-GCN: Spatio-Temporal and Self-Adaptive Graph Convolutional Networks for Single and Multi-View EEG-Based Emotion Recognition," *IEEE Signal Processing Letters,* vol. 29, pp. 1574-1578, June 2022.

[19] G. Zhang, V. Davoodnia and A. Etemad, "PARSE: Pairwise Alignment of Representations in Semi-Supervised EEG Learning for Emotion Recognition," *IEEE Transactions on Affective Computing,* vol. 13, no. 4, pp. 2185-2200, 2022.

[20] R.-N. Duan, J.-Y. Zhu and B.-L. Lu, "Differential Entropy Feature for EEG-based Emotion Classification," *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER),* pp. 81-84, 2013.

[21] W.-L. Zheng and B.-L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Transactions on Autonomous Mental Development,* vol. 7, no. 3, pp. 162-175, 2015.

[22] Q. Wang, O. Sourina and M. K. Nguyen, "Fractal dimension based neurofeedback in serious games," *The Visual Computer,* vol. 27, no. 4, pp. 299-309, Feb. 2011.

[23] S. Spasic, S. Kesic, A. Kalauzi and J. Saponjic, "Different anesthesia in rat induces distinct inter-structure brain dynamic detected by Higuchi Fractal Dimension," *Fractals,* vol. 19, no. 1, pp. 113-123, Mar. 2011.

[24] L. Bhole and M. Ingle, "EEG-based emotion recognition for improving healthcare services," *Cognitive Sensors,* vol. 2, Apr. 2023.

[25] V. M. Joshi and R. B. Ghongade, "EEG Based Emotion Investigation from Various Brain Region Using Deep Learning Algorithm," *ICDSMLA 2020: Proceedings of the 2nd International Conference on Data Science, Machine Learning and Applications,* pp. 395-402, 2022.

[26] L.-C. Shi, Y.-Y. Jiao and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* p. 6627–6630, 2013.

[27] N.-D. Mai, B.-G. Lee and W.-Y. Chung, "Affective Computing on Machine Learning-Based Emotion Recognition Using a Self-Made EEG Device," *Sensors,* vol. 21, no. 15, 2021.

[28] H. Aldawsari, S. Al-Ahmadi and F. Muhammad, "Optimizing 1D-CNN-Based Emotion Recognition Process through Channel and Feature Selection from EEG Signals," *Diagnostics,* vol. 13, no. 16, 2023.

[29] W.-L. Zheng, J.-Y. Zhu and B.-L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," *IEEE Transactions on Affective Computing,* vol. 10, no. 3, pp. 417-429, 2019.

[30] 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," *Biomedical Signal Processing and Control,* vol. 58, 2020.

[31] J. Liu, Y. Zhao, H. Wu and D. Jiang, "Positional-Spectral-Temporal Attention in 3D Convolutional Neural Networks for EEG Emotion Recognition," in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Tokyo, Japan, 2021.

[32] K. Nigam and R. Ghani, "Analyzing the effectiveness and applicability of co-training," *Proceedings of the Ninth International Conference on Information Knowledge Management,* pp. 86-93, 2000.

[33] S. Abney, "Bootstrapping," *Proceedings of the 40th Annual Meeting,* pp. 360-367, 2002.