DeepEnsemble: A Novel Brain Wave Classification in MI-BCI using Ensemble of Deep Learners

Arshad Mehtiyev*, Aziz Al-Najjar*, Hamidreza Sadreazami†, and Marzieh Amini⋆
*System and Computer Engineering Department, Carleton University, Ottawa, Canada
†Bioengineering Department, McGill University, Montreal, Canada
⋆School of Information Technology, Carleton University, Ottawa, Canada

2023 IEEE International Conference on Consumer Electronics (ICCE) | 978-1-6654-9130-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCE56470.2023.10043385

***Abstract*—Brain computer interfaces (BCIs) are rapidly growing in neurorehabilitation to realize communication for consumers with disabilities. This work proposes a classification approach for the motor imagery BCI electroencephalogram (EEG) signals using the recent advances in deep learning methods. The proposed DeepEnsemble method combines different deep learning models in an ensemble learning with soft voting. More specifically, multi-layer perceptron, vision transformer, convolutional neural network and its integration with distributed gradient boosting methods are explored separately and combined in an ensemble learning model. The proposed method is evaluated on a publicly available dataset. It is shown that the proposed ensemble learning method significantly improves the classification accuracy of the EEG signals across various subjects as compared to some of the existing methods.**

***Index Terms*—EEG signal, convolutional neural network, vision transformer model, ensemble learning, common spatial patterns**

I. Introduction

With the advancement of today's technology, having a reliable brain-computer interface (BCI) can not be less than a necessity for patients with neuromuscular disorders [1], [2]. Analyzing the electroencephalogram (EEG) signals and using them as an input to the computer could help individuals unable to speak or who have lost control of their limbs to live once again a comfortable life, where they can communicate, operate and translate these signals into applicable control commands. In addition, EEG signals can be classified for non-medical consumer applications, such as training, education, gaming, and entertainment [3]. However, EEG signals tend to have a non-stationary and distinctive nature. For example, as the human brain is a complex nonlinear system where all body functions are connected, EEG signals of right-hand movements could be mixed with blinking or even breathing signals [4]. This complexity makes it challenging for a simple device to translate the brain signal accurately.

Common spatial patterns (CSP) is a prevalent method to classify EEG signals. [5] was the first to propose the use of CSP to decompose the raw EEG signals. CSP was aimed to maximize the difference between different muscle movements, which later became the foundation of many other CSP-related methods. Later, [6] introduced the regularized CSP (R-CSP) to lower the estimation variance and bias, allowing the algorithm to train on smaller training samples with similar results to the CSP. However, both [5] and [6] were not concerned with the spatial information of EEG electrodes. To solve this issue, [7] suggested using the same R-CSP with a feature associations modeling matrix (MSRCSP) further to smooth the standard CSP with a less training time. Similarly, [8]–[11] continued using modifications of CSP, and combinations with various filters increased the results on average by 10-14% compared to the standard CSP approach in [5]. More recently, [12] used a spectral-temporal CSP to process the EEG signals with a capsule network. The spectral-temporal CSP extracted features while preserved the signals' time resolution.

The implementation of deep learning techniques has given rise to promising results in the EEG signal classification. For instance, convolutinoal neural network (CNN) [13]–[16] and recurrent neural network [17] are commonly used for classifying EEG signal showing superior results as compared to the traditional classifiers. In addition, [18] proposed a hybrid model consisting of CNN with an evolutionary algorithm to classify brainwave signals, where a single discrete wavelet transforms was utilized instead of se Also,veral pooling layers to extract features. [19] was the first to use transformers in EEG signals recognition and obtained comparable results.

In this paper, we propose an ensemble learning method for classification of EEG signals in motor imagery BCI. Inspired by the modified CSP method presented in [12], we first apply CSP to process the EEG signals and extract more distinguishable features. We then train four deep learning models including multi-layer perceptron (MLP), CNN, CNN integrated with distributed gradient boosting and vision transformer (ViT). The models are combined in the proposed DeepEnsemble model and a final classification decision is made by using a soft voting strategy.

II. METHODOLOGY

In this work, we propose a deep-learning based classification algorithm for the EEG signals in motor imagery BCI. The block diagram of the proposed method is shown in Fig.1.

A. Data Pre-processing

The publicly available BCI competition III dataset (IVa) [20] is used to evaluate the performance of the proposed method and to compare it with the results of previous works. The dataset consists of EEG signals of five subjects, referred to as

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

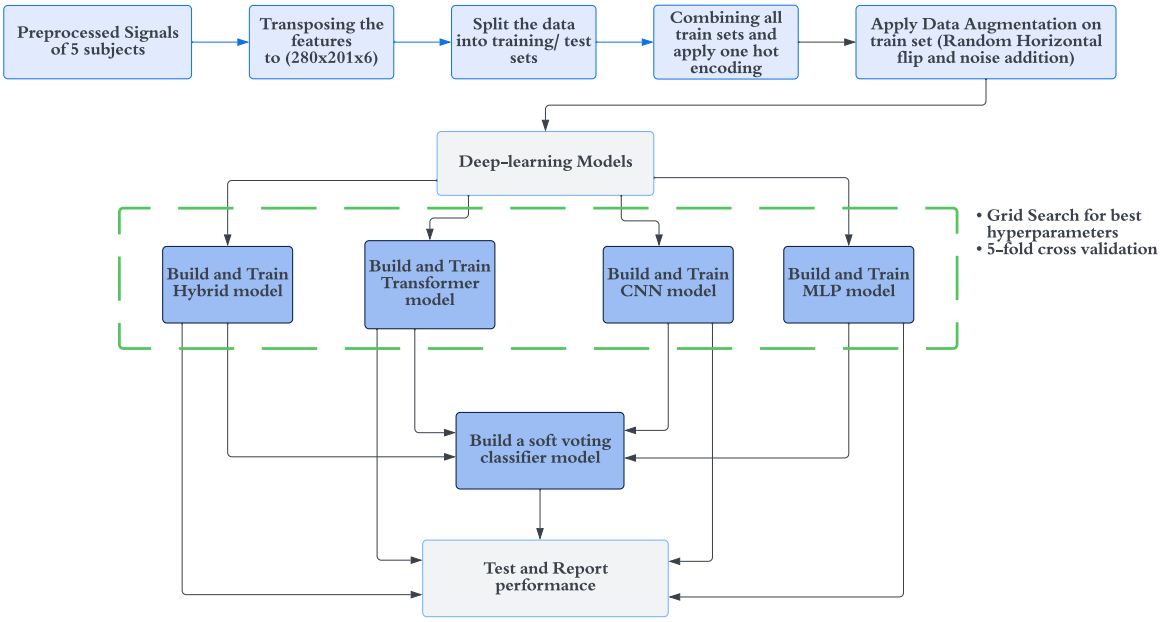


Fig. 1. Block diagram of the proposed DeepEnsemble method.

TABLE I
EEG TRAINING AND TEST DATA SPLIT FOR DIFFERENT SUBJECTS.

Subjects	AA	AL	AV	AW	AY
Trials	280	280	280	280	280
Training set (Tr)	168	224	84	56	28
Test set (Ts)	112	56	196	224	252

AA, AL, AV, AW and AY, while performing limb movements. More specifically, 118 electrodes were used to record 280 trials of right hand and right foot movements with a sample rate of 1000 Hz from each subject. The initial recorded data was filtered between 0.5 and 100 Hz, then a 50 Hz notch filter to suppress line noise and remove unwanted signals, such as blinking and muscle artifacts. Then, following the steps from [12] and [21] to further preprocess the data, the EEG signals went through a 5th-order Butterworth bandpass filter (7-30) Hz range. Also, a weighted moving average filter with a window size of 10 was used to smoothen the signal. CSP is applied to extract the most discriminative patterns in the EEG signals. More specifically, we employ a time-dependant regularized CSP proposed in [12]. This implementation projected spatially-filtered EEG signals onto a 2D representation, which was realized by preserving the information in both the frequency and time domains. This resulted in samples with dimension of 6x201 for each subject in each trial. After preprocessing 280 trials were split into training and test sets separately for each subject, as given in Table I.

First, data were transposed from the shape of (6,201,280) to (280,201,6). This approach allowed changing the orientation of the 3D data matrix so that the data splitting could be done row-wise, where each row represents each trial with 201 feature columns and a depth of 6 EEG channels. Then stratified split was applied to split the dataset for training and testing as per the given ratio in Table I. After the split, data augmentation such as horizontal flip and Gaussian noise addition was applied. This process increased the training set size and helped to avoid over-fitting. As a final pre-processing step, the standard scaler function was devoted to standardizing augmented features as given by

$$z = (x - u)/s, \tag{1}$$

where x is a given sample, u is the mean of the training sample, and s is the standard deviation

B. Multi-layer perceptron model

The first model to be trained for the EEG signals classification is a fully-connected MLP model. This model consists of an input layer to flatten 201x6 features, i.e., 1206 nodes, followed by a network of 20 hidden layers, each containing 100 neuron [22], and an output layer with sigmoid activation. It is noted that the rectified linear unit (ReLU) [23] was used as an activation function.

C. Convolutinoal neural network model

The processed EEG signal has a 2D shape, allowing the use of CNNs. Since the processed EEG data is not an image type as it does not have color depth, an additional single depth dimension is considered in the data before feeding it to the CNN model.

Our CNN model architecture has three 2D convolutional block (Conv2D), a max-pooling layer, a flatten, two fully-connected, and a binary output layer stacked in sequence. The first Conv2D block has 32 filters with a 3x3 window size. The following Conv2D block consists of 64 filters with a 1x1

window size, and the final Conv2D block contains 128 filters with a 3x3 window size. The max-pooling layer has a 1x1 pool size. The initial Dense layer has 128 nodes, and the second has 64 nodes. A dropout is applied after max-pooling and Dense layers with the rates of 0.25, 0.5, and 0.5, respectively. All the layers have a ReLU activation function except for the binary output layer, which has a SoftMax activation function [24]. The architecture of the CNN model is shown in Fig. 2.

D. Hybrid model- CNN integrated with XGBoost

The hybrid model combines the CNN model discussed in Section II.C with the XGBoost algorithm. XGBoost is an algorithm based on scalability and distributed gradient boosted trees [25], and has shown promising results in several applications. In view of this, a combination of CNN with XGBoost is considered in this work. To implement the hybrid model, the processed EEG signals are passed through the trained CNN model and the output is chosen from the first Dense layer. This output is considered an input to XGBClassifier by providing the same label set. This way, the XGBoost classifier is trained. The architecture of the hybrid model is shown in Fig. 3.

E. Vision transformer model

The self-attention-based transformer proved to be promising in several applications [26]. Inspired by the work in [27], we propose a new classifier for the brain activity recognition based on ViT. First, the processed EEG signals are converted into three-channel images to be compatible with the ViT model. Fig. 4(a) shows an image representation of the pre-processed right-hand movement EEG signal, where different colors represent the value of the signal at different points in time. These images are then reshaped into a sequence of flattened 2D patches. The total number of patches N, is defined as

$$N = \text{ image resolution}/(\text{single patch size})^2 \qquad (2)$$

The number of patches is the input sequence length for the transformer, as shown in Fig. 4(c). The patches are then flattened and mapped into a latent vector of size D using a trainable linear projection. This projection step is referred to as patch embeddings. In addition, learnable embedding is appended to the sequence of patches to help represent the signal after the transformer encoder. Before sending the embedding sequence to the encoder, position embeddings are

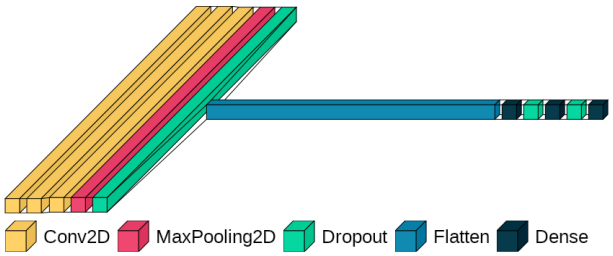


Fig. 2. CNN model architecture

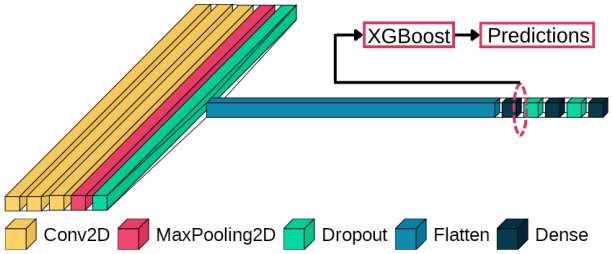


Fig. 3. Hybrid model architecture

added to retain the positional information. The transformer encoder consists of alternating layers of multi-headed self-attention and MLP blocks, layer normalization before every block, and residual connections after each block.

F. Ensemble learning with soft voting (DeepEnsemble)

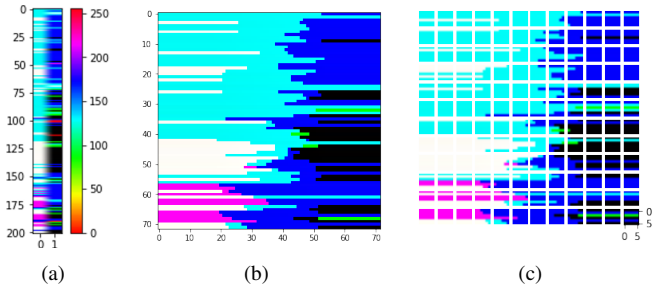
Ensemble learning with soft voting is implemented to combine the probabilistic predictions of the classifiers, discussed in Section II(B-E), to make a final prediction. The soft voting algorithm is considered to sum all the predicted class probabilities. Then the class with the highest score is assigned either 0 or 1.

III. Experimental Results and Discussion

To evaluate the performance of the proposed method, the following metrics are calculated for each model and each subject: accuracy, precision, recall, AUC and F1 score. The results were obtained based on 5-fold cross-validated data sets with stratification, and the average of performance metrics was recorded.

Table II gives the performance result comparison of each model on subject AA. Table III lists the classification accuracy of each model for different subjects.

It is seen from these tables that in some cases, the MLP model performs remarkably better than the other models with a small dataset. In addition, as the train/test dataset split ratio differs for each subject, ViT delivers better results, when the training dataset is large, as in subjects AA and AL. This also indicates that in such a binary classification problem, when converting EEG signals into 3D representation, the ViT model provides comparable results to CNN without the need for



(a) (b) (c)

Fig. 4. 4(a): Processed EEG signal as an image with the size of 201x2 pixels and 3 color channels. 4(b): Resized image of 4(a) to 72x72. 4(c): Divided image in 4(b) into 144 patches each of size 6x6

TABLE II
PERFORMANCE METRICS COMPARISON OF EACH MODEL ON SUBJECT AA

	MLP	***ViT***	***CNN***	***Hybrid***	***DeepEnsemble***
Accuracy	95.71±1.73	93.75±2.65	93.21±2.37	90.54±0.91	96.07±0.91
Precision	98.12±1.15	95.18±0.98	92.4±3.31	87.96±0.57	96.08±0.67
Recall	93.21±3.07	92.14±4.87	94.29±2.08	93.93±1.82	96.07±1.75
AUC	95.71±1.73	93.75±2.65	93.21±2.37	90.54±0.91	96.07±0.91
F1 Score	95.58±1.85	93.59±2.81	93.31±2.26	90.84±0.95	96.06±0.94

TABLE III
PERFORMANCE ACCURACY COMPARISON OF DIFFERENT METHODS FOR EACH SUBJECT. (VALUES IN BOLD INDICATE THE BEST RESULT, AND IN PARENTHESIS INDICATE THE SECOND BEST RESULT)

	AA	***AL***	***AV***	***AW***	***AY***
MLP	(95.71±1.73)	**95.0±1.75**	**89.08±2.69**	**91.52±0.89**	(95.16±1.11)
ViT	93.75±2.65	93.21±2.08	87.35±1.56	(90.98±1.58)	94.92±0.95
CNN	93.21±2.37	90.0±4.01	86.73±0.91	86.43±3.5	91.43±2.02
Hybrid	90.54±0.91	89.64±2.86	86.02±1.43	85.89±3.0	90.63±1.24
DeepEnsemble	**96.07±0.91**	(94.64±2.53)	(88.78±1.83)	90.62±1.16	**96.03±0.25**

a huge dataset for pre-training and fine-tuning, unlike [27]. It is also observed that as the BCI dataset under study is relatively small, both ViT and CNN show acceptable accuracy but also expose a high standard deviation between trials. This is where combining CNN with XGBoost comes in handy. The scalability and distributed gradient boosted tree features of the XGBoost helped reduce the standard deviation of the CNN model by almost up to 62% for all subjects, except the subject AV. The results in Table II also show that with the optimized hyper-parameters, the MLP outperforms the other models with a mean accuracy of 95.71% on subject AA. ViT is coming second at 93.75%, with the highest standard deviation of 2.65%. CNN is slightly standing behind at 93.21%. Meanwhile, combining the CNN model with XGBoost forming the hybrid model provided the lowest accuracy of 90.54% among the models. However, in comparison, it had the lowest standard deviation of 0.91%.

It is seen from Table III that the MLP, ViT and DeepEnsemble models performed slightly better than CNN and the hybrid model. As DeepEnsemble combines in itself all four models, it reduces the standard deviation of classification in most of the subjects to a great extent, and it provides considerably better result in subject AY, which has the lowest training samples.

The proposed DeepEnsemble model combines all the models in an ensemble learning with soft voting, resulting in a mean accuracy value of 96.07%, which surpasses the MLP model with the help of ViT and CNN, while keeping the standard deviation at 0.91% with the use of the XGBoost. It is noted that the ensemble model utilizes the best features of the used models. It is also noted that the results showed similar patterns between subjects. For illustration, the accuracy comparison of each model and the proposed DeepEnsemble for subject AA is shown in Fig. 5.

Comparing with some of the state-of-the-art methods, results of the proposed DeepEnsemble method on every subject are 5-fold cross-validated with stratification, and the average performance accuracy is listed in Table IV. As can be seen from this table, the proposed method significantly outperforms other existing methods by providing higher classification accuracy across subjects. For instance, subject AA results outperformed the current top result, FBCSP [11], by 6.14%, while it surpassed FBRCSP [9] by 25% for subject AV. Also, for subject AY, it bested EEGCAPS [12] by 12.46%. However, the DeepEnsemble model accuracy for subject AL and AW is slightly less - by 3.57% from SBCSP [8] and by 4.07% from EEGCAPS [12], respectively, but still is among the top three results.

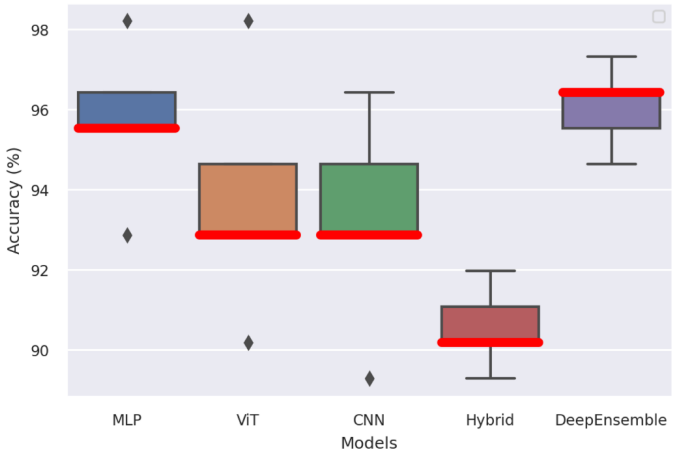


Fig. 5. Box plots for classification accuracy of different methods on subject AA.

TABLE IV
COMPARISON OF THE PROPOSED DEEPENSEMBLE METHOD WITH THE OTHER EXISTING METHODS.(VALUES IN BOLD INDICATE THE BEST RESULT, AND IN PARENTHESIS INDICATE THE SECOND BEST RESULT)

	AA	***AL***	***AV***	***AW***	***AY***
CSP [5]	66.07	96.43	43.47	71.88	49.60
R-CSP [6]	77.7	96.4	58.7	(92.0)	68.3
MSRCSP [7]	69.64	96.43	59.18	71.88	52.78
SBCSP [8]	83.03	**98.21**	52.04	89.05	58.33
FBCSP [11]	(89.93)	96.43	63.26	72.32	54.37
SSRCSP [10]	70.54	96.43	53.57	71.88	74.39
FBRCSP [9]	84.82	96.43	(63.78)	74.55	73.81
LRDS [28]	80.4	94.6	50.0	90.6	83.3
EEGCAPS [12]	85.50	(97.52)	62.15	**94.70**	(83.57)
DeepEnsemble	**96.07**	94.64	**88.78**	90.63	**96.03**

IV. CONCLUSION

This paper has proposed a methodology for classifying the EEG brainwave signals in motor imagery BCI using modified CSP for pre-processing and DeepEnsemble classification method. The proposed DeepEnsemble method has combined different deep learning models including MLP, CNN, CNN integrated with XGBoost and ViT models in an ensemble learning with soft voting. Extensive experiments were performed on a publicly-available dataset. The result have shown that the proposed DeepEnsemble method outperforms the other existing methods by providing higher classification accuracy across subjects. The proposed method can pave the way for future ensemble models and improve the accuracy of the EEG signal classification in the MI-BCI.

REFERENCES

[1] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," *Mayo clinic proceedings*, vol. 87, no. 03, pp. 268–279, 2012.

[2] G. Kalantar, H. Sadreazami, and A. Asif, "Adaptive dimensionality reduction method using graph-based spectral decomposition for motor imagery-based brain-computer interface," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 990–994, 2022.

[3] J. Van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.

[4] W. Klonowski, "Everything you wanted to ask about eeg but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, no. 1, pp. 1–5, 2009.

[5] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[6] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for eeg classification in small-sample setting," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[7] X. Li and H. Wang, "Smooth spatial filter for common spatial patterns," in *International Conference on Neural Information Processing*, pp. 315–322, 2013.

[8] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcsp) for brain-computer interface," in *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, pp. 204–207, 2007.

[9] S.-H. Park and S.-G. Lee, "Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 2977–2983, 2017.

[10] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE world congress on computational intelligence)*, pp. 2390–2397, 2008.

[12] H. Sadreazami and G. D. Mitsis, "Motor task learning in brain computer interfaces using time-dependent regularized common spatial patterns and residual networks," in *2020 18th IEEE International New Circuits and Systems Conference (NEWCAS)*, pp. 190–193, 2020.

[13] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to mi-eeg signal classification for bcis," *Expert Systems with Applications*, vol. 114, pp. 532–542, 2018.

[14] X. Wang, M. Hersche, B. Tömekce, B. Kaya, M. Magno, and L. Benini, "An accurate eegnet-based motor-imagery brain–computer interface for low-power edge computing," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2020.

[15] H. Sadreazami, Y. Taheri, and M. Amini, "Hierarchical spectral-temporal feature learning for motor task recognition in brain computer interfaces," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–5, 2022.

[16] A. Khoyani, H. Kaur, M. Amini, and H. Sadreazami, "Motor imagery brain activity recognition through data augmentation using dc-gans and mu-sigma," in *2022 IEEE Sensors*, pp. 1–4, 2022.

[17] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "Lstm-based eeg classification in motor imagery tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2086–2095, 2018.

[18] Z. R. K. Rostam and S. A. Mahmood, "Classification of brainwave signals based on hybrid deep learning and an evolutionary algorithm," *arXiv preprint arXiv:1912.07361*, 2019.

[19] J. Sun, J. Xie, and H. Zhou, "Eeg classification with transformer-based models," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 92–93, 2021.

[20] B. Blankertz and K.-R. Muller, "Data set iva(motor imagery, small training sets)." BCI competition III, Dec. 12, 2004 [Online].

[21] H. Sadreazami, M. Amini, M. Omair Ahmad, and M. N. S. Swamy, "Eegcaps: Brain activity recognition using modified common spatial patterns and capsule network," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, 2021.

[22] H. Faris, I. Aljarah, and S. Mirjalili, "Training feedforward neural networks using multi-verse optimizer for binary classification problems," *Applied Intelligence*, vol. 45, no. 2, pp. 322–332, 2016.

[23] K. Fukushima, "Cognitron: A self-organizing multilayered neural network," *Biological Cybernetics*, vol. 20, no. 3-4, pp. 121–136, 1975.

[24] J. S. Bridle, "Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters," in *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, NIPS'89, (Cambridge, MA, USA), p. 211–217, MIT Press, 1989.

[25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[28] R. Tomioka and K. Aihara, "Classifying matrices with a spectral regularization," in *2007 24th International Conference on Machine Learning*, pp. 895–902, 2007.