The final accepted and published version of this paper can be accessed at
https://ieeexplore.ieee.org/document/10295823

A Novel Signal-to-Image Conversion Approach with Ensembles of Pretrained CNNs for Motor Imagery EEG Signals

1st Cagatay Murat Yilmaz
Department of Software Engineering
Karadeniz Technical University
Trabzon, Turkey
cmyilmaz@ktu.edu.tr

2nd Bahar Hatipoglu Yilmaz
Department of Computer Engineering
Karadeniz Technical University
Trabzon, Turkey
baharhatipoglu@ktu.edu.tr

3rd Cemal Kose
Department of Computer Engineering
Karadeniz Technical University
Trabzon, Turkey
ckose@ktu.edu.tr

***Abstract*—Deep learning techniques can recognize and learn significant features automatically. The advantages of deep learning should also be used to solve problems in other fields. Therefore, we concentrated on classifying motor imagery EEG signals with deep transfer learning. The present study first proposed an EEG signal-to-image conversion method that uses 2D signal representations. Thanks to this, we employed 2D images as input instead of standard one-dimensional EEG features. Second, we used AlexNet, GoogLeNet, and SqueezeNet pre-trained convolutional neural network frameworks for classification. We also used hard voting and merged the outputs of pre-trained convolutional neural networks. We performed experiments on dataset 2a of BCI Competition IV, achieving 84.18±5.37% classification accuracy and 0.66±0.12 kappa across nine subjects, even with limited training data. The results showed that the proposed signal-to-image conversion method with deep transfer learning has potential, particularly in processing biomedical signals.**

***Index Terms*—motor imagery, brain-computer interface, convolutional neural networks, deep transfer learning, signal-to-image conversion**

I. Introduction

Brain-computer interfaces (BCI) bypass the human neuromuscular and establish a communication medium between the brain and computers. Motor imagery (MI) is a common way of designing BCIs. In these interfaces, actions are performed by stimulating the brain without using the motor system. Although machine learning has produced promising results, many problems, such as session-to-session and subject-to-subject variability, a time-consuming calibration stage, still need to be solved. Deep learning techniques have recently become popular among the most promising candidates for MI-based BCIs. Until recently, most MI-EEG research has been confined to classical feature extraction and classification approaches. Feature extraction approaches include time/frequency/time-frequency domain features, spatial features, adaptive methods, and combinations. Classification methods include nearest neighbor classifiers, Bayesian classifiers, linear discriminant analysis approaches, artificial neural networks, support vector machines, adaptive classifiers, Riemann approaches, classifier combinations, and ensemble learning. Due to its superior automated feature extraction and classification skills compared to machine learning, deep learning has gained much popularity. Nowadays, they are among the most popular research themes in academia and the industry. It has attracted much attention, and many novel deep-learning practices have been put forward. They have been used to classify different data types and demonstrated great success. Regarding deep learning architectures, convolutional neural networks (CNNs) are among the most famous.

Deep learning methods can classify MI-EEG signals with 1D signals or feature inputs. However, these techniques typically use 2D image/feature inputs or processes in two dimensions instead of 1D time series MI-EEG signals or features. Therefore, MI-EEG signals should be converted to two-dimensional formats, such as images or features. It is also true for other types of biomedical signals. One option for classifying MI-EEG signals with CNNs is to convert the signals into images. In the literature, time-frequency representation methods like continuous wavelet transform (CWT) and short-time Fourier transform (STFT) have generally been utilized to transform MI-EEG signals into 2-D scalograms or spectrograms. For example, [1] constructed 2D MI-EEG images of adaptively reconstructed signals using CWT-based time-frequency maps. These images were inputted into AlexNet and LeNet CNN architectures. The proposed approaches achieved promising outcomes compared to the previous studies and showed the potential of MI-EEG task recognition using image classification. Deep CNN classification accuracy suffers due to the frequency bands' time-varying nature. To overcome these problems, [2] converted EEG signals into 2D images with STFT and then built a model with transfer learning and EfficientNetB0. The proposed approach successfully classified left and right-hand MI tasks. In some studies conducted with STFT and CWT, images generated from different electrodes or frequency bands carrying MI task information were combined. For instance, [3] used a quadratic time-frequency distribution and transformed MI-EEG signals into time-frequency images (TFIs). These images represent the energy distributions of MI-EEG signals in terms of time, frequency, and electrode. They also used a new CNN architecture to extract TFIs'

features automatically. The proposed method achieved good results in decoding the MI tasks. Besides the appropriate input to CNNs, another critical problem is that deep learning models need a lot of data. However, in MI-EEG studies, it is challenging to collect data. This difficulty increases even more with complex mental tasks. Transferring deep learning models can provide excellent solutions in areas where training data is scarce, and costs are high. In transfer learning, the model is not created from scratch. Instead, pre-prepared features and learning models are transferred from one problem domain to another. It also automatically extracts richer and more expressive features during training [4].

Only a few methods have been proposed in the state-of-the-art to convert MI-EEG signals to images or features. They are limited to basic techniques like STFT and CWT, which convert EEG signals to time-frequency representations. In the present study, we aimed to propose a novel signal-to-image conversion method using the quasi-probabilistic distribution models proposed in [5], which can be an alternative to STFT and CWT. The present study also addressed the data scarcity problem of MI-EEG datasets. We used AlexNet, GoogLeNet, and SqueezeNet pre-trained CNN frameworks to classify a new collection of MI-EEG images. We also employed hard voting and merged the outputs of these pre-trained CNNs. The experimental results on classifying two-class left and right-hand MIs demonstrated that the proposed signal-to-image conversion method with deep transfer learning has good capacity. The paper is organized as follows. The proposed signal-to-image-transformation approach is given after introducing the dataset in Section 2. Section 3 presents the experiments, the results obtained, and comparisons with the literature. In the last section, conclusions and recommendations are given.

II. Material and Methods

A. Experimental Data

In the literature, many data sets have been proposed for MI-EEG. Most of them are benchmark datasets and contain the following MI tasks: hand (left/right-hand), foot (left foot, right foot, both feet), tongue, and fist (both fists, opening and closing left or right fist). In the present study, we performed experiments on BCI Competition IV Data Set 2a, including challenges like eye movement artifacts. This dataset includes four categories of MI tasks recorded from nine subjects. These are motor imagery tasks of the left/right hand, tongue, and feet. Training and test sets were created with two sessions recorded on different days. Recordings were taken from 22 electrodes and sampled at 250 Hz. They were then bandpass filtered between 0.5–100 Hz and notch-filtered at 50 Hz [6] [7]. This study investigated two-class left and right-hand MI task classifications. The EOG artifacts were reduced, and trials containing human expert scoring artifacts were not used. The number of trials is shown in Table I. The number of MI-EEG trials in each category is almost the same.

TABLE I: The distribution of trials for left- and right-hand MI tasks for Data Set 2a

	Training Session			Test Session		
Subject	LH	RH	Sum	LH	RH	Sum
A01	69	69	138	71	70	141
A02	67	69	136	71	71	142
A03	69	68	137	67	70	137
A04	62	67	129	59	57	116
A05	63	66	129	70	65	135
A06	56	57	113	53	55	108
A07	67	66	133	71	69	140
A08	66	66	132	66	68	134
A09	53	63	116	65	65	130

B. Methods

1) Signal-To-Image Conversion: This paper proposed a novel signal-to-image conversion method based on quasi-probabilistic distribution models of [5] and 2-D modeling images of [8]. The proposed method used a feature representation approach that utilizes 2-D feature matrices as in [5]. The representativeness of EEG signals varies from subject to subject, and different results may be obtained in particular channels and frequency bands. This study combined images from different frequency bands and channels to examine the effects of different factors. For a one-channel MI-EEG signal, we converted EEG signals to images as follows:

Step 1: Calculate $M \times N$-sized 2D feature matrix.

Step 2: Apply a zero-mean Gaussian kernel to smooth.

Step 3: Scale the content of the matrix to the full range of the 'jet'color map.

When applying this method, we detected transition points with a value of 0.05 and employed a ($w = [21 \times 21], \omega = 4.0$) Gaussian Kernel. The 2-D representation and hence the images were created in 50×50 size. C3, CZ, and C4 electrode locations were used in experiments because they are in MI-related brain regions. Furthermore, to capture event-related desynchronization and synchronization, μ and β bands were employed. The images to be given as input to the network were created as follows. Firstly, EEG signals were filtered using a fifth-order Butterworth with bands of 6$\sim$13 Hz, 17$\sim$30 Hz, and 4$\sim$38 Hz. They include μ and β and a more comprehensive range of frequencies. After that, the EEG signals were converted into 100×100-sized sub-images using the method mentioned above (steps 1 to 3). Finally, these sub-images were merged to create the fused final image, as shown in Fig. 1. In this phase, we first horizontally combined sub-images of 6$\sim$13 Hz, 17$\sim$30 Hz, and 4$\sim$38 Hz frequency bands, respectively, and constructed 300×100-sized images for a one-channel EEG signal. After that, we vertically combined images of channels while preserving neighboring information and formed a 300×300-sized final image. When merging, sub-images related to frequency bands were resized to the same size to ensure similar effects for all bands.

2) Hard Voting: It is one of the simplest examples of ensemble learning. It determines the final classification label

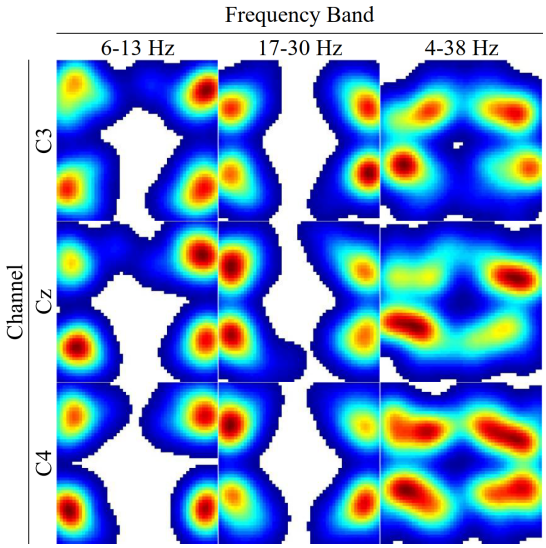

Fig. 1: The conversion process of EEG signals to images

$\hat{y}$ of the whole system using the most frequent label output $O_k(x)$ of m sub-classifiers (voters) because each sub-classifier has the same weight on the final decision [9] [10]. For example, for a binary classification problem, let $m = 3$ sub-classifiers produce $O_1(x) \leftarrow class\,0$, $O_1(x) \leftarrow class\,1$ and $O_1(x) \leftarrow class\,0$ output labels, respectively [9]. Since two sub-classifiers are labeled as class 0 and one classifier is labeled as class 1, a hard voter classifies it as $\hat{y} = 0$ according to the most frequent label.

$$\hat{y} = mode\left\{O_1(x), O_2(x), \ldots, O_m(x)\right\} \tag{1}$$

III. Experiments and Results

The original training (first session) and test (second session) sets were merged in experiments. All EEG signals were then converted into images. All images were randomly divided into 90% training and 10% test sets for each subject. It was accomplished five times, and the average of the five runs was used as the final result. Experiments were conducted only to classify left and right-hand MIs. Results were obtained for each subject separately. Classification results were obtained for accuracy and kappa metrics. Accuracy and kappa are the most commonly used values in the literature for comparison in studies on this dataset.

In the study, some crucial factors in the performance evaluation and their values are as follows. We extracted 2.5 s-long (corresponding to 625 samples) signals for each EEG trail between 0.5 s and 3.0 s after the cue onset. C3, Cz, and C4 channel signals were filtered in 6∼13 Hz, 17∼30 Hz, and 4∼38 Hz bands. The proposed signal-to-image conversion was applied to the time-series EEG data. The same parameter values were used for all subjects to generate the 2-D images (see Sec. II-B1 for details). Also, no data augmentation was employed during training. Sample 2D input images generated from left and right hand MI tasks are shown in Fig. 2, respectively.

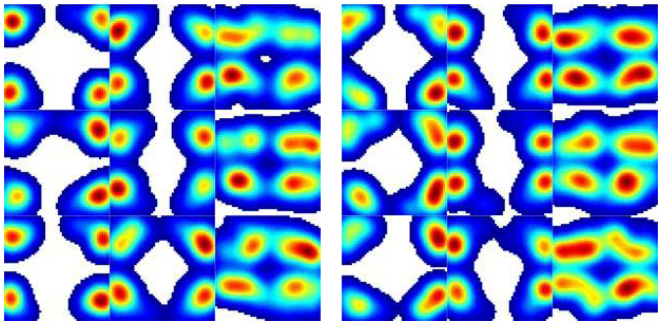

Fig. 2: Sample images for (a) left and (b) right hand signals

This study used AlexNet, GoogLeNet, and SqueezeNet pre-trained CNNs to classify a new collection of MI-EEG images. As in [4], we adopted the network without layer modifications, fine-tuned the final layers of CNN architectures, and employed transfer learning to train the pre-trained CNNs to classify the newly proposed MI-EEG images. SqueezeNet, GoogLeNet, and AlexNet were pre-trained on the ImageNet data set containing 1000 categories. It helped pre-trained networks to learn rich feature representations. The learned features and models were transferred directly to the target CNN for MI classification. We also replaced the last layers with new ones to learn features specific to the MI tasks, as in [11]. An early stopping strategy was used in the training to avoid overfitting problems. In fine-tuning, 1e-3, 1e-4, and 1e-5 were determined as initial learning rates. Adam, RMSProp, and stochastic gradient descent with momentum optimizers were used to minimize the error. The highest number of epochs was selected as 50. 2 to 32 (exponents of 2) were used as mini-batch sizes. Different parameter combinations in different folds and subjects yielded the most successful results. Therefore, the best performance values were used when presenting the results.

Table II shows the average performance results across five runs. The averages and standard deviations across subjects are also given in the last rows. Considering the average performance values across all subjects, AlexNet achieved the best result with 79.77±7.31% classification accuracy (CA) and kappa of 0.57±0.16. Other CNN architectures obtained slightly worse results on average across subjects. GoogLeNet showed CA of 79.27±7.05% with a kappa of 0.56±0.16, and SqueezeNet 77.72±6.69% with a kappa of 0.53±0.15. In the second experiment, we used AlexNet, GoogLeNet, and SqueezeNet pre-trained CNNs as sub-classifiers and employed hard voting that uses the most frequent sub-classifier outputs. Hard voting achieved an average CA of 84.18±5.37% with a kappa of 0.66±0.12 across subjects. It showed CA from 77.78% (for A07) to 94.81% (for A03) and kappa from 0.53 (for A01) to 0.89 (for A03) for subjects. As can be seen from the results across subjects, the hard voting strategy enhanced the performance of individual pre-trained CNNs. It indicates that the hard voting strategy provided both a 4.41%

TABLE II: Results for pre-trained CNN-based methods on Dataset 2a for the classification left/right hand MIs

Metric	Approach	A01	A02	A03	A04	A05	A06	A07	A08	A09	Avg±Std
Accuracy (%)	AlexNet	73.33	74.81	91.85	74.17	73.85	81.82	76.30	78.46	**93.33**	79.77±7.31
	GoogLeNet	72.59	78.52	89.63	78.33	73.85	80.00	72.59	74.62	**93.33**	79.27±7.05
	SqueezeNet	77.78	72.59	89.63	71.67	72.31	79.09	70.37	77.69	88.33	77.72±6.69
	Hard Voting	**80.00**	**82.96**	**94.81**	**86.67**	**80.77**	**84.55**	**77.78**	**79.23**	90.83	**84.18±5.37**
Kappa	AlexNet	0.40	0.45	0.83	0.43	0.46	0.63	0.51	0.53	**0.87**	0.57±0.16
	GoogLeNet	0.33	0.55	0.78	0.52	0.46	0.59	0.44	0.47	**0.87**	0.56±0.16
	SqueezeNet	0.53	0.41	0.78	0.36	0.43	0.58	0.39	0.52	0.76	0.53±0.15
	Hard Voting	**0.53**	**0.62**	**0.89**	**0.66**	**0.60**	**0.69**	**0.54**	**0.56**	0.82	**0.66±0.12**

CA improvement and a 1.94% standard deviation reduction compared to the best of Alexnet. A similar improvement was observed in kappa. The results revealed that hard voting enhanced classification performance.

The conclusions for Data Set 2a were also compared with the state-of-the-art, as shown in Table III. They have mainly used accuracy and kappa metrics. In the literature, due to the paucity of data, deep learning studies like [12] generally merged training and test sets and evaluated them with k-fold CV. On the other hand, classical classification methods like [13] and [14] have used the first session's data for the training and the second for the evaluation. In [12], the authors studied bi-class (left/right hand) and three-class (left/right hand and foot) MI task classification problems. They developed a CNN architecture with spatial brain neural patterns' enhanced interpretations. For this purpose, they extracted 2D features using Power Spectral Density and CWT. They employed topographic interpolation for projecting the multi-channel data and preserved the spatial interpretations. Besides, they eliminated learned weights, including a spatial dropping algorithm. In this paper, for a bi-class problem, CWT without spatial dropping + CNN approach achieved an average CA of 87.6±5.7% with a kappa of 0.75 using a 10-fold CV. For a three-class problem, CWT with spatial dropping + CNN approach showed an average CA of 71.2±7.0% with a kappa of 0.56 using a 10-fold CV. These best results were achieved with a thresholding strategy combined with Continuous Wavelet Transform. It improved CNN's interpretability and classification accuracy at the same time.

The human brain produces non-stationary EEG oscillations, and the EEG-based BCIs' input distributions vary. It is a significant problem for the development of online adaptive systems. Due to this, [14] investigated a novel integration of unsupervised adaptive ensemble learning (CSE-UAEL) and covariate shift estimation to solve for MI-EEG. They first detected CSP features' covariate shifts using an exponentially weighted moving average model. After that, they constructed a classifier ensemble and updated it for streaming input data distribution changes. The proposed CSE-UAEL + C-3 achieved an average CA of 80.86±11.44% in the passive scheme and 81.48±11.33% in the active scheme. In this paper, other baseline methods like AdaBoost and TotalBoost obtained average CA rates between 71.53±11.76% and 75.15±13.44%. In another study, [13] designed a monolithic and modular network for multi-class MI-EEG problems. This paper employed Bayesian optimization, processed the data in the spectral and spatial domains, and extracted features using a Discriminative Filter Bank Common Spatial Pattern variant. The most successful frequency bands are selected iteratively, and over the testing session, the monolithic network showed an average CA of 78.41±5.91% with 0.59±0.11 kappa. This paper obtained the best results using the modular network with an average CA of 80.03±6.52% and a kappa score of 0.61±0.12.

Compared to the previous studies, the hard voting of the pre-trained CNNs approach increased the best average classification accuracies for two subjects: 85.4% of CWT+CNN [12] to 86.67% for A04, 79.22% of modular network [13] to 80.77% for A05. A similar trend was seen for kappa values. The proposed method showed significant improvements for two subjects: 0.69 of modular network [13] to 0.89 for A03 (a considerable improvement) and 0.68 of CWT+CNN [12] to 0.82 for A09 (a significant improvement). In the state of the art, CWT+CNN [12] achieved the best results for other subjects and across subjects. This study obtained an average of 3.42% better results across subjects. However, the standard deviation of our research is 0.33% lower. Compared to other methods of modular network [13] and CSE-UAEL+C-3 [14], more successful average results of 2.7% and 4.15% were obtained, respectively. Furthermore, lower standard deviations of 1.15% and 5.96% were acquired. [12] improved the average results with a kappa of 0.09. On the other hand, our method outperformed [13] with an average kappa of 0.05. The results are significant because they were obtained with limited training data and without data augmentation. Moreover, the signal-to-image transform method is easy to compute.

IV. Conclusion

In the present study, we proposed a signal-to-image conversion approach and classified the converted images using various pre-trained CNNs. Additionally, we combined and voted the results of pre-trained CNNs. Successful results were obtained on both a subject-by-subject basis and across subjects.

TABLE III: Comparison with studies on Dataset 2a to classify left/right hand MI tasks

	Accuracy				Kappa		
	CWT+CNN [12]	Modular Network [13]	CSE-UAEL+C-3 [14]	Hard Voting of pre-trained CNNs	CWT+CNN [12]	Modular Network [13]	Hard Voting of pre-trained CNNs
A01	83.4	84.91	**91.67**	80.00	**0.88**	0.68	0.53
A02	**83.8**	66.38	63.89	82.96	**0.73**	0.36	0.62
A03	**95**	84.74	94.44	94.81	0.67	0.69	**0.89**
A04	85.4	81.36	72.22	**86.67**	**0.73**	0.62	0.66
A05	79.1	79.22	77.08	**80.77**	**0.90**	0.6	0.60
A06	**86.7**	70.67	75.69	84.55	**0.71**	0.45	0.69
A07	**86.4**	86.12	73.61	77.78	0.58	**0.71**	0.54
A08	94	83.81	**94.44**	79.23	**0.90**	0.72	0.56
A09	**94.8**	83.04	90.28	90.83	0.68	0.66	**0.82**
Avg±Std	**87.6±5.7**	80.03±6.52	81.48±11.33	84.18±5.37	**0.75±0.11**	0.61±0.12	0.66±0.12

The proposed signal-to-image transformation approach and voting of the outputs of different CNNs played a crucial role in these results. The results were obtained with limited labeled data and without augmentation. It is crucial when considering the data scarcity problem of MI-EEG systems. The models can be further improved with improvements in signal-to-image conversion and deep neural networks and by improving the data set with data augmentation. The computational cost of the proposed signal-to-image transformation approach is lower than time-frequency domain methods such as STFT and CWT. Therefore, it also has the potential for real-time MI-EEG applications. In the present study, we used transfer learning to fine-tune the CNN models, and the best results in diverse subjects and folds were obtained with different hyperparameter values. Therefore, such deep learning problems must also be solved. In the future, we aim to improve the signal-to-image transformation method, perform data augmentation suitable for MI-EEG, and create CNN models by training from scratch. Thus the classification performance can be improved. We also seek to apply the proposed signal-to-image conversion method to other time series problems.

ACKNOWLEDGMENT

This research was supported by the Turkish Scientific and Research Council (TUBITAK) under project number 119E397. The authors thank the Knowledge Discovery Institute, Graz University of Technology, for sharing the dataset. The authors also thank the Council of Higher Education of Turkey.

REFERENCES

[1] N. Alizadeh, S. Afrakhteh, and M. R. Mosavi, "Deep cnn-based classification of motor imagery tasks from eeg signals using 2d wavelet transformed images of adaptively reconstructed signals from mvmd decomposed modes," *International Journal of Imaging Systems and Technology*, 2023 May 22 2023.

[2] M. S. Ali, A. Hassan, A. Rahim, M. H. Ashraf, A. Rahim, and S. Saghir, "Motor imagery eeg classification using fine-tuned deep convolutional efficientnetb0 model," in *2023 3rd International Conference on Artificial Intelligence (ICAI)*, 2023, pp. 1–6.

[3] R. Alazrai, M. Abuhijleh, H. Alwanni, and M. I. Daoud, "A deep learning framework for decoding motor imagery tasks of the same hand using eeg signals," *IEEE Access*, vol. 7, pp. 109 612–109 627, 2019.

[4] M. Wei, R. Yang, and M. Huang, "Motor imagery eeg signal classification based on deep transfer learning," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, ser. IEEE International Symposium on Computer-Based Medical Systems, 2021, pp. 85–90.

[5] C. M. Yilmaz, C. Kose, and B. Hatipoglu, "A quasi-probabilistic distribution model for eeg signal classification by using 2-d signal representation," *Computer Methods And Programs in Biomedicine*, vol. 162, pp. 187–196, Aug 2018.

[6] "Bci competition iv - data sets 2a." [Online]. Available: https://www.bbci.de/competition/iv/#datasets

[7] M. Tangermann, K.-R. Mueller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schoegl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv," *Frontiers in Neuroscience*, vol. 6, 2012.

[8] C. M. Yilmaz, "Classification of eeg-based motor imagery tasks using 2-d features and quasi-probabilistic distribution models," Doctoral dissertation),, The Graduate School of Natural and Applied Sciences, Karadeniz Technical University, Trabzon, Turkey, 2021.

[9] S. Raschka, "Mlxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack," *Journal of Open Source Software*, vol. 3, no. 24, p. 638, 2018. [Online]. Available: https://doi.org/10.21105/joss.00638

[10] R. Hidayati, K. Kanamori, L. Feng, and H. Ohwada, "Implementing majority voting rule to classify corporate value based on environmental efforts," in *Data Mining and Big Data, DMBD 2016*, ser. Lecture Notes in Computer Science, vol. 9714, 2016, pp. 59–66.

[11] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional neural network based approach towards motor imagery tasks eeg signals classification," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4494–4500, Jun 15 2019.

[12] D. F. Collazos-Huertas, A. M. Álvarez-Meza, C. D. Acosta-Medina, G. A. C. Duque, and G. Castellanos-Domínguez, "Cnn-based framework using spatial dropping for enhanced interpretation of neural activity in motor imagery classification," *Brain Informatics*, vol. 7, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221472594

[13] B. E. Olivas-Padilla and M. Chacon-Murguia, I, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Applied Soft Computing*, vol. 75, pp. 461–472, FEB 2019.

[14] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface," *Neurocomputing*, vol. 343, no. SI, pp. 154–166, May 28 2019.