Machine Learning for Multi-Action Classification of Lower Limbs for BCI

Zilu Wang
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
zw20774@essex.ac.uk

Jichun Li
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
jl20340@essex.ac.uk

Ian Daly
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
i.daly@essex.ac.uk

*Junhua Li**
School of Computer Science and Electronic Engineering
University of Essex
Colchester, UK
Wuyi University
Jiangmen, China
junhua.li@essex.ac.uk

***Abstract*—Over the past two decades, significant progress has been made in brain-computer interfaces (BCIs), devices which enable direct communications between human brains and external devices. One of the prevalent control paradigms is motor imagery-based BCI (MI-BCI), by which users imagine specific actions to express their intentions. Left-hand and right-hand motor imageries are frequently used in the MI-BCI. If a third class is needed, the imagination of both feet is usually added. However, it is relatively rare to separate feet into left lower limb and right limb in MI-BCI systems. In addition, previous studies have demonstrated that real movements can be distinguished from one another via processing the electroencephalogram (EEG). Similarly, motor imagery (MI) and movement observations (MO) can also be distinguished from one another. However, classification of left lower limb actions and right lower limb actions between MI, Real Movement (RM), and MO actions, has not been thoroughly explored. To address these questions, we performed a comprehensive experiment to collect EEG under six actions (i.e., Left-MI, Right-MI, Left-RM, Right-RM, Left-MO, and Right-MO) and used three models (convolutional neural network [CNN], support vector machine [SVM], and a K-Nearest Neighbours [KNN]) to classify these actions. Our CNN achieved the highest performance (37.77%) in the classification of six actions. Although the performance of SVM (37.21%) and KNN (25.26%) was worse, it is still better than the chance level (16.67%). Our results suggest that it is possible to distinguish between these six lower limb actions. This study has implications for developing multi-class BCI systems and promoting the research of multiple-action classification.**

Keywords— Brain-Computer Interface; EEG; Lower Limbs; Motor Imagery; Movement Observation; Real Movement

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) provide a direct communication pathway between human brains and external devices without the involvement of peripheral nerves and muscles [1]. One widely used BCI control paradigm is the motor imagery (MI) paradigm, in which people just imagine movements of a body part in order to control the BCI. In other words, users express their intentions by MI, and these intentions can be decoded based on electroencephalography (EEG) signals [2]. Amongst other uses, MI is a promising technology for healthy people to learn new motor skills in sports [3], and it is also extremely useful in the rehabilitation training for the motor recovery of paralyzed people [4].

The brain regions activated by MI are similar to those activated by real executive movements [5]. Therefore, for patients who are difficult to exercise during rehabilitation, they can imagine movements instead of real movements to stimulate the damaged motor network and restore the lost motor functions [6]. In addition to MI, movement observation (MO) was also found to positively affect neurological rehabilitation training. Babiloni et al. [7] found that similar brain areas are activated during MO and execution in experiments on animals and humans. Specifically, they reported that there is "a functional equivalence" or "shared motor representations" between the cortical processes underlying MO and execution [7].

Similar brain regions are activated by MI and MO, but differences between them have been observed in previous studies. Berends et al. [8] explored the difference between MI-and-MO and MO-only in the sensorimotor areas (e.g., at EEG electrodes C3, C4) and the central parietal cortex (e.g., Pz). The results show that during MI-and-MO, the modulation of EEG rhythms was stronger than that in the case of MI-only in the theta (4~8Hz), alpha (8~12Hz), and beta (13~25Hz) frequency bands. Grafton et al. [9] used positron emission tomography imaging of cerebral blood flow to localize brain regions associated with hand grasping movement during MI and MO. This study demonstrated that there were differential activations of areas for hand grasping movement in MI and MO. In another study, McFarland et al. [10] investigated similarities and differences between the effects of RM and MI on the mu and beta rhythms during right-hand and left-hand movements. The results indicated that MI and RM cause modulations in mu and beta rhythm amplitudes and that decoding these changes is crucial for aiding communication through BCI systems. Babiloni et al. [11] extracted the alpha (the frequency range about 10 Hz) and beta (the frequency range about 20 Hz) rhythms to compute event-related desynchronization/synchronization (ERS/ERD) in association with MO and RM of unilateral left and right aimless finger movement. Their results

suggested that no matter which side of the body the event occurred, bilateral central ERD was significantly stronger in the RM than in MO action ($p<0.01$).

Collectively, these studies demonstrate that MI, MO, and RM lead to different brain activities, although the activated brain regions (i.e., sensorimotor area) are similar. To date, the majority of studies focus on upper limbs actions. This might be because it is more difficult to explore lower limbs compared to upper limbs. One of the difficulties is that the brain regions supervising the lower limbs are closer to the central sulcus, making the signals generated by these brain regions harder to differentiate via non-invasive neuroimaging. For example, Batula et al. [12] explored cortical activation differences between MI and RM in the tasks of the finger- or toe-tapping actions (left finger-tapping, right finger-tapping, left toe-tapping, and right toe-tapping) using functional near-infrared spectroscopy (fNIRS). The results showed that it was more difficult to differentiate activation patterns between left toe-tapping and right toe-tapping compared to that of the finger-tapping actions. To the best of our knowledge, no research has been carried out to classify multiple actions (including RM, MI, and VM) of separate lower limbs based on EEG signals. We will address this in our study.

Machine learning is usually used to automatically achieve recognition or classification and has been applied in many fields such as image analysis [13], natural language processing [14], motion control [15], and BCI [16]–[20]. Traditional machine learning approaches, such as support vector machine (SVM) [16] k-nearest neighbours (KNN) [17], and Bayesian classifiers [18], have been employed in the classification of EEG signals. In recent years, Deep Learning (DL) has become increasingly popular. In 2015, Li et al. employed the Lomb–Scargle periodogram to estimate the spectral power and denoising autoencoder based neural network to perform binary MI classification [19]. The classification performance of this deep learning model (i.e., denoising autoencoder) is compared with the SVM model [19]. The performance of the denoising autoencoder was not absolutely and significantly better than that of SVM. Following this work, more deep learning models have been applied to classify EEG signals. The convolutional neural network (CNN) classifier is one of the most frequently used classification methods for MI tasks in the field of DL [20]. For example, Zhang et al. [21] used the CNN model to classify MI tasks (i.e., left and right-hand) based on the features extracted by the short-time Fourier transform (STFT) and obtained excellent classification performance.

In this study, we designed an experiment to induce brain activities associated with lower limbs corresponding to three conditions (i.e., MI, MO, and RM) and addressed whether actions of the left and right lower limbs can be classified based on EEG signals. Both traditional machine learning methods and deep learning models were used for the classification and compared in terms of classification performance.

II. Materials and Methods

A. Experiment Design

In the experiment, participants were asked to perform tasks according to cues presented on a computer monitor. There were six sessions. Each session consisted of 72 trials. Every trial started with a fixation cross, shown at the centre of the screen. The duration the fixation cross was shown for was randomised from 1.5 seconds to 2.5 seconds (the mean was two seconds). This was followed by a condition cue, which was a triangle filled with a colour. Red, green, and blue colours were used to represent the three task conditions (RM, MI, and MO). The colours did not consistently correspond to the conditions and were randomised across participants (e.g., the red colour representing motor imagery for one participant might be used to represent real movement for another participant). One second after the condition cue, an arrow appeared above the condition cue. The arrow was a direction cue, indicating either the left direction or the right direction. Both cues remained on the screen for six seconds. Once the direction cue appeared, participants needed to keep performing the required task according to the cues of condition and direction until the cues disappear. After that, the fixation cross appeared on the screen again, and the next trial started. There were six classes in total, combining the three conditions and the two directions. All trials were presented in random order. The timing protocol of the trials is shown in Fig.

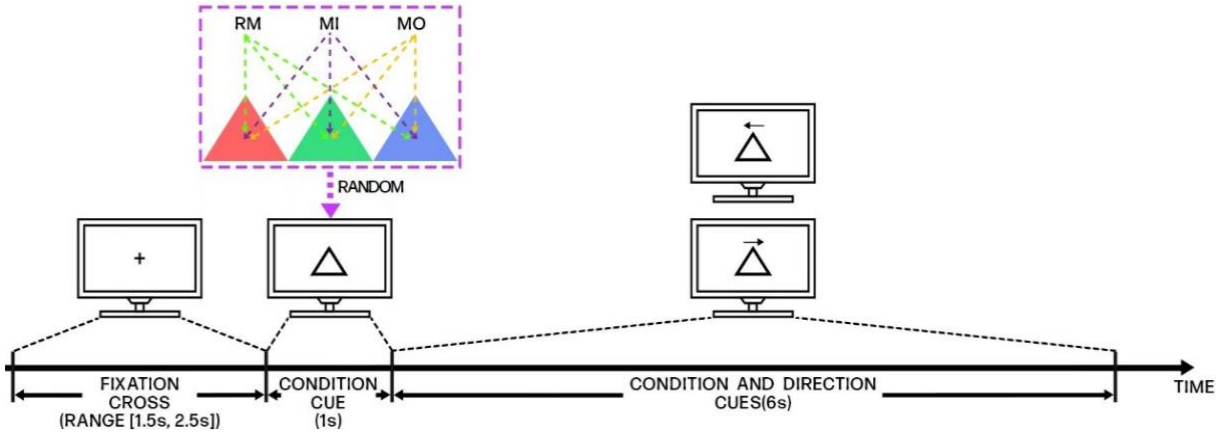

Fig. 1. The timing of a trial. At the biginning of each trial, a fixation cross appears on the screen for a period varying from 1.5 seconds to 2.5 seconds. At the stage of the condition cue, a triangle appears on the screen for one second. Different colours correspond to RM, MI, and OM conditions. Colours were randomised to represent the conditions. Last, during the condition and direction cues stage, a triangle and a direction arrow appear on the screen.

B. Data Acquisition

Fifteen healthy participants participated in the experiment. Sixty-two electrodes were used to record EEG signals according to the international 10-20 standard system. Two pairs of electrodes were used to record vertical and horizontal EOG signals. All signals were recorded at a sampling rate of 500 Hz and referenced to the average of both mastoids. The impedance was maintained below 15 kilo-ohms. We recruited fifteen healthy participants for this study. The study was reviewed and approved by the Institutional Review Board of the National University of Singapore, and the Humanities, Science and Health, or Social Science Ethics Sub-Committee at the University of Essex.

C. Data Processing

The EEG and EOG signals were down-sampled to 250 Hz. Abnormal channels were visually detected and replaced by interpolation using the surrounding four channels. Eye movement-related artefacts were mitigated in EEG signals using the adaptive filtering method [22]. Then, a band-pass filter (0.5 Hz ~ 45 Hz) was applied to the EEG signals. Subsequently, EEG signals were divided into eight-second-long segments, from one second before the onset of the condition cue to the end of the condition and direction cues. Independent component analysis (infomax) was used to decompose these EEG segments into components. An experienced research assistant manually identified the components relevant to artefacts. These components were removed, and the remaining components were used to reconstruct clean EEG segments.

D. Feature Extraction and Classification

Short-time Fourier transform (STFT) is used as a feature extraction method, which is an effective method in time-frequency analysis for non-stationary and non-linear signals [23], [24]. It was selected to measure the power spectral density (PSD) of the EEG signals for feature extraction in this study. The portion of EEG data corresponding to task implementation is used for feature extraction, resulting in six-second long data segments. The time series of each channel $X_i(n)$ ($i = 1, 2, 3, \cdots 62$) in a segment was converted into spectral features by STFT as follows:

$$F(m,\omega) = \sum_{n=-\infty}^{\infty} X_i(n)\, W(n - mR) e^{-j\omega n} \quad (1)$$

Where $W(n)$ is Hamming function with the length of 250. R is set to 125. The PSD $P(m,\omega)$ is calculated from $F(m,\omega)$ by:

$$P(m,\omega) = |F\ (m,\omega)|^2 \quad (2)$$

$P(m,\omega)$) is a two-dimensional matrix with the size of 125×11 (the direct current component is excluded). We selected a range of frequencies between 8 and 30 Hz to obtain a new PSD matrix of 23×11 as features of each channel. The matrix (23×11) was reshaped into a vector with a length of 253. After obtaining vectors for each channel, we assembled these vectors into a two-dimensional matrix (62×253) as features.

Five-fold cross-validation was applied to evaluate the performance of three models (CNN, KNN, SVM). CNN model was customized to perform the classification. As the parameter settings influence the classification performance, we tuned them to maximize the performance of the CNN model. All parameters were divided into two groups: sensitive parameters and non-sensitive parameters. According to previous studies [25]–[28], the following parameters were sensitive: the decay of the learning rate, the kernel size, the number of filters (in the first and second convolutional layers), the dropout rate, the size of maxpooling, and the number of neurons in the fully-connected layer. A grid search was used to find out optimal values for these sensitive parameters using training set. The optimal values of these sensitive parameters are listed in Table I. The values of non-sensitive parameters were determined based on our previous empirical experience.

TABLE I. LIST OF OPTIMAL SENSITIVE PARAMETERS IDENTIFIED VIA OUR GRID SEARCH.

Parameters	Values
Decay of learning rate	0.99
Kernel size of convolution	(8,5)
Number of filters (first convolutional layer)	32
Number of filters (second convolutional layer)	64
Dropout rate	0.5
Kernel size of maxpooling	(2,2)
Number of neurons in the fully-connected layer	128

The model architecture is illustrated in Fig. 2. A CNN model was constructed: The first layer is the input layer; the second layer is a convolutional layer with a kernel size of 8×5, the number of filters is 32; the third is a max pooling layer with a kernel size of 2×2; then there is another convolutional layer with a kernel size of 8×5, the number of filters is 64; this is followed by another maxpooling layer with a kernel size of 2×2; the output of the maxpooling is flattened and mapped into 128 neurons (fully-connected layer), which is followed by 6 neurons. These correspond to 6 categories: MI of left lower limb (Left-MI), MI of right lower limb (Right-MI), RM of left lower limb (Left-RM), RM of right lower limb (Right-RM), MO of left lower limb (Left-MO), and MO of right lower limb (Right-MO). After each convolutional layer, ReLU (Rectified Linear Unit) is used as an activation function. Batch normalization and dropout are applied to reduce the risk of overfitting in this model.

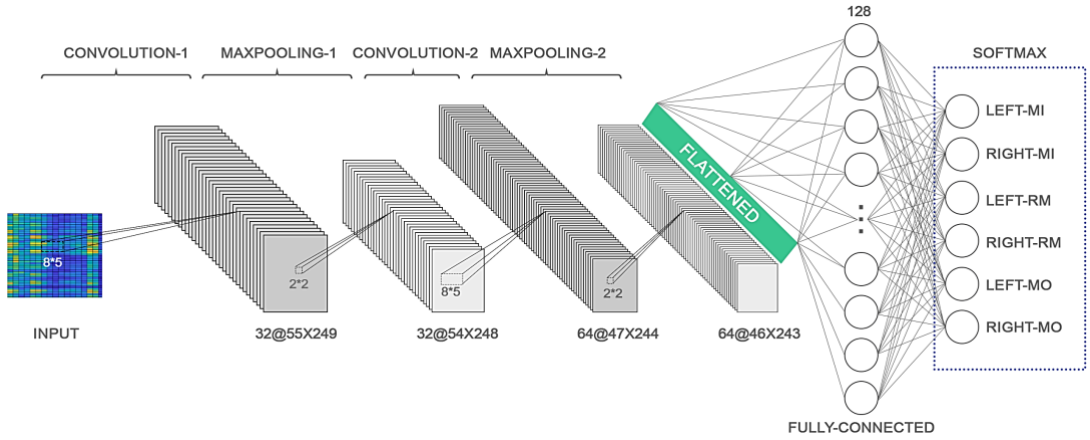

Fig. 2. The architecture of the CNN model. The model consists of two convolutional layers, two maxpooling layers, and two fully-connected layers. Adam was used as the optimization algorithm for this study.

In this study, we also used KNN and SVM models to classify the data. The KNN algorithm is one of the simplest object classification algorithms in the machine learning [29]. A new sample is classified based on the distances between this sample and the training samples. The class that the majority of k nearest training samples is assigned to the new sample and is considered as the classification outcome. The performance of the KNN classifier depends on the distance metric and the value of the neighbourhood parameter k. Therefore, the k value is tuned ($k = 3, 5, 10$). We found that the highest classification performance was obtained when $k = 5$. Thus, we set the k value as 5. Euclidean distance was used to measure the distances between samples. The SVM model is one of the most widely used classification techniques for high-dimensional vectors [30]. It aims to find a hyper-plane in the feature space to maximize the distance between the hyper-plane and the nearest data points of each class. The type of kernel function is important for the architecture of the SVM. We used the linear kernel function in the SVM model to classify the data. We chose this kernel because the dimension of training data was high and the amount of calculation was large.

III. Results and Discussion

We compared the CNN model with the SVM and KNN models in terms of classification performance. the CNN models achieved the highest average accuracy of 37.77%, whereas the SVM, and KNN models had accuracies of 37.21% and 25.26% (see Table II). The classification performance was analysed first via a one-way analysis of variance (ANOVA). ANOVA showed a significant difference in the performance among the three classifiers ($F_{(2,42)} = 10.56,\ p < 0.001$). Furthermore, we measured the statistical significance of the differences in classification performance between classifiers using two-tailed paired t-test to assess whether the differences in classification accuracies are significant. Our results show that the CNN significantly outperformed the KNN ($t_{14} = 5.93,\ p < 10^{-4}$). The mean classification accuracy of CNN is higher than SVM, but there is no significant difference ($t_{14} = 0.29,\ p > 0.05$). The performance of the SVM classifier is also significantly better than the KNN ($t_{14} = 14.07,\ p < 10^{-7}$). Fig. 3. also shows that all methods performed better than the chance level (16.67%). This result shows that six different lower limb-related action tasks are separable.

TABLE II. Performance Metrics for our Proposed CNN, SVM and KNN across all Participants

Participants	Accuracy (%)		
	CNN	SVM	KNN
P1	56.17	51.16	37.03
P2	36.62	33.8	20.62
P3	39.12	29.19	23.17
P4	54.19	50.54	39.33
P5	32.88	31.24	19.88
P6	26.89	30.11	17.82
P7	38.45	35.42	31.07
P8	20.62	38.84	26.12
P9	18.29	32.44	21.79
P10	40.5	37.5	23.4
P11	41.89	38.43	23.81
P12	30.13	32.65	22.91
P13	36.79	31.01	17.42
P14	48.17	45.4	30.58
P15	45.81	37.21	25.26
Mean±STD	37.77±10.59	37.21 ± 6.84	25.26±6.30

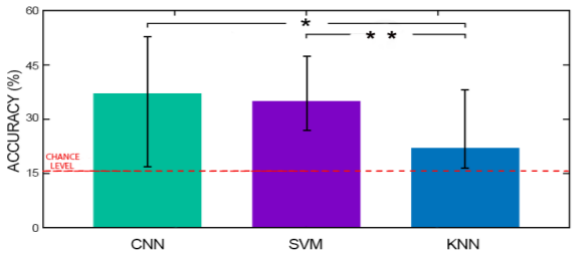

Fig. 3. Classification accuracy comparison among three methods: CNN, SVM, and KNN (* indicates $p < 10^{-4}$, ** indicates $p < 10^{-7}$). The dotted line in the figure indicates the chance level (16.67%).

To know the classification accuracy of each action more clearly, we calculated the confusion matrix for the CNN classifier, which is the classifier with the best classification performance. Fig. 4. shows the average of all participants' confusion matrices. The diagonal elements of the matrix represent the proportion of correct classification, and the non-diagonal elements represent the respective proportions that were incorrectly classified into the other categories. The sum of each column of the matrix is 100%. It can be seen

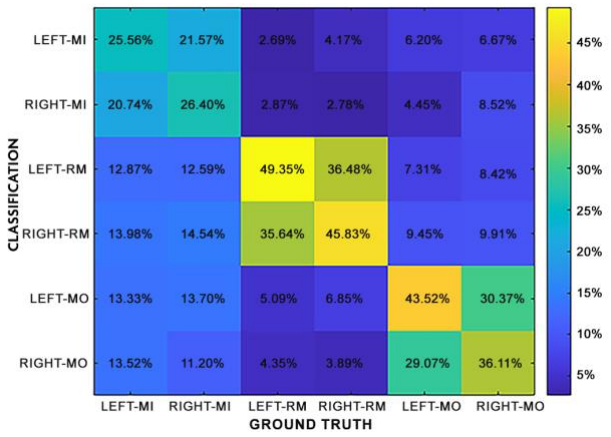

Fig. 4. The confusion matrix shows the average classification accuracies across 15 participants.

that the accuracies in the diagonal are generally high, in which RM is the action with the highest classification accuracy.

In order to intuitively observe the differences between lower limb multi-action tasks, nine channels with high correlation with the movement of left and right feet were selected for analysis: F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4. One of the 15 participants was randomly selected to illustrate the average power spectral densities, which are shown in Fig. 5. It can be seen that EEG power begins to increase after around 8 Hz and converges after around 13 Hz. This exhibits that the larger differences between tasks appear in the frequency band of 8 Hz ~ 13 Hz compared to the other bands. This finding suggests that EEG signals in this band may have potential value in classifying lower limb actions. In addition, the plots show that between 8 and 13 Hz, the difference between the left lower limb and right lower limb in the same condition (MI, RM, or MO) is less than that between conditions.

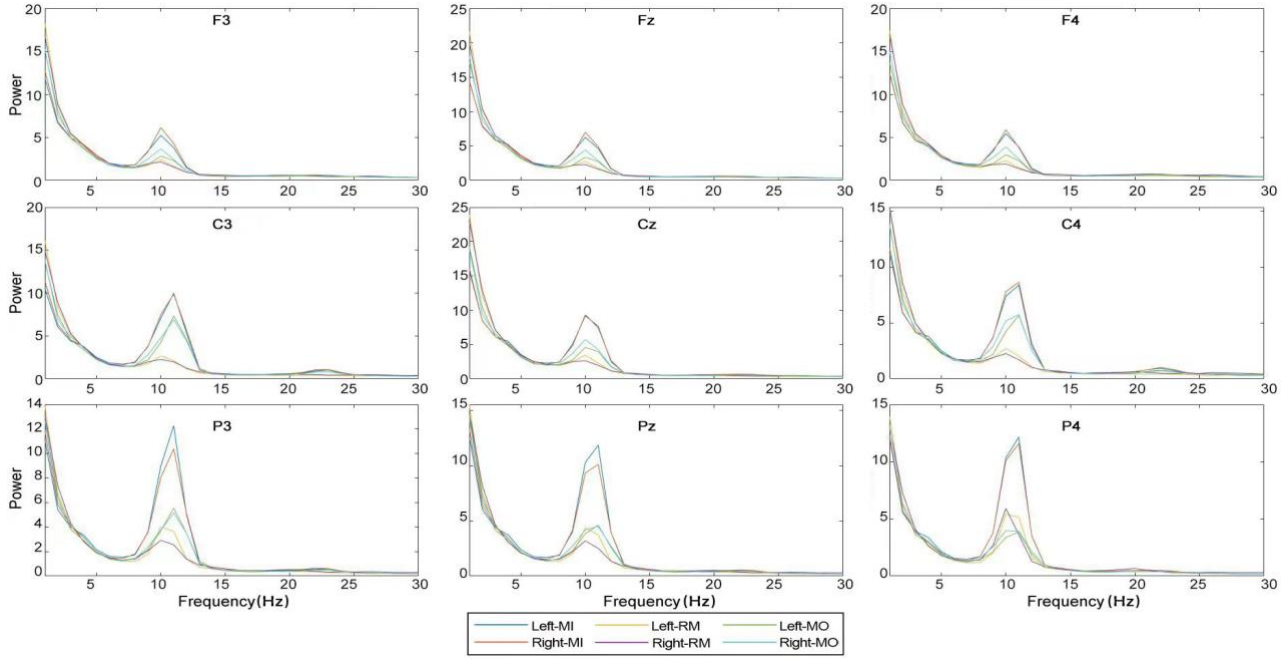

Fig. 5. Average power spectral densities of six actions for nine representative channels. Each curve in the plot corresponds to an action performed by participants in the experiment. The higher overlap of the curves, the less separable between actions.

IV. Conclusion

In this paper, we designed an experiment to collect EEG under six actions (i.e., Left-MI, Right-MI, Left-RM, Right-RM, Left-MO, and Right-MO). The data were classified using three classification models (e.g., CNN, SVM, and KNN). The classification results show that the actions of the left and right lower limbs can be classified based on EEG signals. The classification performance of the three models was compared. It was found that the CNN model outperformed both SVM and KNN models. In addition, we also found that the multi-action of lower limbs was more distinguishable in the frequency band of 8 ~ 13 Hz than in the other frequency bands. In the near future, we plan to increase the number of participants to enhance the confidence of the results derived based on the study. We also have planned to develop sophisticated models to improve classification performance.

References

[1] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," Electroencephalogr. Clin. Neurophysiol. vol. 78, no. 3, pp. 252–259, Mar. 1991.

[2] M. Jeannerod, "Mental imagery in the motor contex,"Neuropsychologia, vol. 33, no. I1, pp. 1419–1432, 1995.

[3] S. M. Murphy, "Imagery interventions in sport," Med. Sci. Sports Exerc. vol. 26, no. 4, pp. 486–494, 1994.

[4] N. Sharma, V. M. Pomeroy, and J. C. Baron, "Motor imagery: A backdoor to the motor system after stroke?," Stroke, vol. 37, no. 7, pp. 1941–1952, Jul. 2006.

[5] S. Kraeutner, A. Gionfriddo, T. Bardouille, and S. Boe, "Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations," Brain Res. vol. 1588, pp. 81–91, Nov. 2014.

[6] S. J.- Neuroreport and undefined 2000, "Imagining the impossible: intact motor representations in hemiplegics," NeuroReport, vol. 11, no. 4, pp. 729-732, 2000.

[7] F. Babiloni et al., "Recognition of imagined hand movements with low resolution surface laplacian and linear classifiers," Med. Eng. Phys., vol. 23, on. 5, pp.323-328, 2001.

[8] H. I. Berends, R. Wolkorte, M. Ijzerman, and M. J. A. M. Van Putten, "Differential cortical activation during observation and observation-and-imagination," Exp. Brain Res. vol. 229, pp. 337–345, 2013.

[9] S. T. Grafton, M. A. Arbib, L. Fadiga, and G. Rizzolatti, "Localization of grasp representations in humans by positron emission tomography," Exp. Brain Res. vol. 112, pp. 103–111, 1996.

[10] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," Brain Topogr. vol. 12, no. 3, pp. 177–186, 2000.

[11] C. Babiloni et al., "Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study," Neuroimage, vol. 17, no. 2, pp. 559-572, 2002.

[12] A. M. Batula, J. A. Mark, Y. E. Kim, and H. Ayaz, "Comparison of brain activation during motor imagery and motor movement using fNIRS," Comput. Intell. Neurosci., vol. 2017, 2017.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, pp. 770-778, 2016.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding," Proceedings of NAACL-HLT, Minneapolis, pp. 4171-4186, 2019.

[15] L. Xiao, J. Dai, R. Lu, S. Li, J. Li, and S. Wang, "Design and comprehensive analysis of a noise-tolerant ZNN model with limited-time convergence for time-dependent nonlinear minimization," IEEE Trans. Neural Netw. Learn. Syst. vol. 31, no. 12, pp. 5339–5348, 2020.

[16] T. Zhang, W. C.-I. T. on N. S. and, and undefined 2016, "LMD based features for the automatic seizure detection of EEG signals using SVM," IEEE Trans. Neural Syst. Rehabil. Eng. vol. 25, no. 8, pp. 1100-1108, 2017.

[17] B. Hu, X. Li, S. Sun, M. R.-I. transactions on, and undefined 2016, "Attention recognition in EEG-based affective learning research using CFS+ KNN algorithm," IEEE/ACM Trans. Comput. Biol. Bioinform, vol. 15, no. 1, pp. 38-45, 2016.

[18] Y. Zhang et al., "Sparse Bayesian classification of EEG for brain–computer interface," IEEE Trans. Neural Netw. Learn. Syst. vol. 27, no. 11, pp. 2256-2267, 2016.

[19] J. Li, Z. Struzik, L. Zhang, and A. Cichocki, "Feature learning from incomplete EEG with denoising autoencoder," Neurocomputing, vol. 165, pp. 23-31, 2015.

[20] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review," Neural Comput. Appl. pp. 1-42, 2021.

[21] J. Zhang, C. Yan, and X. Gong, "Deep convolutional neural network for decoding motor imagery based brain computer interface," IEEE internatiional conference on signal processing, communications and computing (ICSPCC), Xiamen, 2017.

[22] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering," Med. Biol. Eng. Comput. vol. 42, no. 3, pp. 407–412, 2004.

[23] T. H. Shovon, Z. A. Nazi, S. Dash, M. F. Hossain, "Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT," The international conference on advances in electrical engineering (ICAEE), Dhaka, 2019.

[24] D. Trad, T. Al-ani, E. Monacelli, M. Jemni. "Nonlinear and nonstationary framework for feature extraction and classification of motor imagery," 2011 IEEE international conference on rehabilitation robotics, Zurich, 2011.

[25] N. Lu, T. Yin, X. Jing. "Deep learning solutions for motor imagery classification: A Comparison Study," 8th international winter conference on brain computer interface (BCI), South Korea, 2020.

[26] B. Abibullaev, I. Dolzhikova, and A. Zollanvari, "A Brute-Force CNN Model Selection for Accurate Classification of Sensorimotor Rhythms in BCIs," IEEE Access, vol. 8, pp. 101014–101023, 2020.

[27] D. Feng, L. Chen, and P. Chen. "Intention Recognition of Upper Limb Movement on Electroencephalogram Signal Based on CSP-CNN," 5th international conference on robotics and automation sciences (ICRAS), Wuhan, 2021.

[28] W. Samuel et al., "A channel selection approach based on convolutional neural network for multi-channel EEG motor imagery decoding," IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, 2019.

[29] K. Fukunaga, "Introduction to statistical pattern recognition," Elsevier 2013.

[30] T. Kayikcioglu, and O. Aydemir "A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data," Pattern Recognit. Lett. vol. 31, no. 11, pp. 1207-1215, 2010.