Channel Selection and Classification of Electroencephalography based Motor Imagery Signals using Machine Learning

Rab Nawaz
James Watt School of Engineering,
University of Glasgow,
Glasgow G12 8QQ, UK
Rab.Nawaz@glasgow.ac.uk

Pot Kheng Wang
Department of Electronic Engineering
Universiti Tunku Abdul Rahman
Kampar 31900, Malaysia
potkheng98@1utar.my

Muhammad Burhan Khan
Department of Electrical Engineering
National University of Computer and Emerging Sciences
Karachi, Pakistan
burhan.khan@nu.edu.pk

Humaira Nisar
Department of Electronic Engineering
Universiti Tunku Abdul Rahman
Kampar 31900, Malaysia
humaira@utar.edu.my

Abstract—**Brain Computer Interface (BCI) is a topic of research for many years. It is beneficial for people with limited mobility, in addition BCI is also applied in other fields such as entertainment, security, education. Electroencephalography (EEG) is the most utilized brain signals in BCI systems. The accurate classification of EEG signals is highly desirable for BCI applications. In this study motor imagery (MI) EEG signals are classified into left- and right-hand movements. Publicly available dataset from PhysioNet project is utilized in this study. Four kinds of features including band power, approximate entropy, statistical features, and wavelet features were extracted from the data. These features were used individually as well in different combinations to test their reliability in classifying the MI tasks. Three different classifiers, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) were used to classify the data. The capability of CNN to classify the raw signal instead of extracted features from the data is also investigated. Additionally, the performance of the CNN model with the reduced number of EEG channels is tested. The best result was obtained using CNN as a classifier with statistical features, with a classification accuracy of 94.28%. The effect of different combinations of features and the reduced number of EEG channels are discussed in the paper.**

Keywords—Motor imagery, EEG, BCI, KNN, SVM, CNN

I. INTRODUCTION

The brain is one of the most complex structures in the universe. Electroencephalogram (EEG) is one of the recording techniques that acquire the electrical activity emanating from the brain on the scalp and is commonly used in brain computer interface (BCI). BCI builds communication between the brain and an external device such as a computer or machine [1]. A wide range of devices are implemented with BCI, but most significant are those that are used in neuro-prosthetics application that helps to restore or facilitate the patients [2]. With the help of BCI, patients with paralysis, locked-in syndrome, advanced Amyotrophic Lateral Sclerosis (ALS), spinal cord injury or any diseases that result in losing the ability to move, can perform the intended task with minimal support and hence it improves their quality of life.

Motor imagery (MI) is mental rehearsal of execution of motor task without actual movement [3] and is described as a mental imagination of movements. When human beings perform MI task, the brain motor areas will generate activity known as sensorimotor rhythms (SMR). Many research studies are conducted to classify MI EEG signals. Different features and classification algorithms have been proposed in the past. Isa et. al. used machine learning to perform classification of MI movements in their study [4]. They used Fast Fourier Transform-Linear Discriminant Analysis (FFT-LDA) technique to extract features from EEG and classify it using different classifiers including K-Nearest Neighbors (KNN), Decision Tree, Logistic Regression, support vector machine (SVM), and Naïve Bayes. In classifying left-hand and right-hand motions, the highest accuracies were obtained with Naïve Bayes (79.23%), and SVM (78.89%). Wave Common Spatial Pattern (WaveCSP) that uses Wavelet Transform (WT) and Common Spatial Pattern (CSP) filtering techniques is proposed in [5]. It utilized SVM, LDA, and KNN classifiers. The classification accuracy obtained was 63.4% for SVM and LDA, and 63.5% for the KNN classifier. In another study, a deep learning model using Convolutional Neural Network (CNN), was adopted to classify 5 different MI tasks including hand close, hand open, wrist pronation, wrist supination, and do nothing from amputees [6]. In this study, 64 channels were used to train the CNN model, later the number of channels was reduced to 8 based on the trained model. Mean classification accuracy of 99.7% was obtained for 64 electrodes, while 91.5% of accuracy was obtained from 8 selected channels. A recent study is done by Shajil et al. on classifying spatially filtered MI EEG signals using CNN [7]. Raw EEG was acquired using 16 channels and filtered between 1 and 100Hz. Five channels (C3, C1, CZ, C2, C4) were selected to study the performance of minimal EEG electrodes. CSP is used to improve the information contained in the channels by increasing the spatial resolution. A single convolutional layer CNN system, with 16 filters of kernel size 3 was proposed. In two-class (left hand and right hand) and four-class (left hand, right hand, two hands, and feet) classification, they obtained 95.18% and 87.37% accuracy, respectively. Another study reported by [8] used five EEG signal features from 64 electrodes from 50 participants, i.e., Band Power, Approximate Entropy, statistical features, wavelet-based features, and CSP. For classification, Decision Tree (DT), Random Forest (RF), SVM, KNN, and Artificial Neural Network (ANN) were used. These methods were tested on Physionet MI database for four different MI tasks. They reported an accuracy of 98.53% for left and right hands MI tasks with Approximate Entropy feature and SVM classifier.

To improve the performance of BCI systems, different EEG features and algorithms for classification have been studied [9]. Unquestionably, BCI devices appear to be one of the best solutions for those suffering from motor disabilities. The effectiveness, user-friendliness, functionality, comfort, and mobility are the key features of a daily-use device. Although substantial BCI research has been carried out and a lot of new technologies regarding BCI have been invented throughout these years, BCI products are still rarely seen in t market due to their relatively high cost for common consumers. Therefore, reducing the number of electrodes is a great way to increase the cost-effectiveness of the device, while providing the same performance by removing the irrelevant and redundant channels. The reduction of BCI electrodes has been reported in [10], with a very high classification accuracy however the researcher's used their own dataset and hence the results could not be compared with other state of art algorithms. Hence to classify the MI based cognitive tasks with the reduced number of EEG channels the current work is focused on:

- Exploring EEG features that are related to MI to classifying EEG into different classes
- Identifying optimum EEG channels (electrodes) to reduce the number of EEG channels without loss of accuracy

II. METHODOLOGY

A. Data Acquisition and Pre-processing

The EEG MI dataset created by Schalk et al. [11] is used in the current study. Data were collected from 109 participants that consists of over 1500 EEG recordings. 64-channels BCI2000 system with 160 sampling rate was used for data acquisition. Each participant performed 14 experimental runs which are summarized in Table 1 [8,11]. The EEG data were free from artifacts (non-brain activities such as ocular artifacts, muscle artifacts) [12] as the BCI2000 system has built-in artifact removal algorithms. Previous studies suggest that events related to MI are mainly manifested in the mu (8–12 Hz) and beta rhythms (12–30 Hz). Therefore, the signal is bandpass filtered between 0.5 and 30Hz. Short segments ranging from 1.5 sec before and 2.5 sec after the trial onset, resulting in 4-sec long epochs are created from the continuous data. Each epoch was baseline corrected by subtracting the mean power in the pre-onset signal from the whole epoch.

Table 1 Experimental runs for each subject

Tasks	Duration	No. of runs
Baseline, eyes open	1 minutes	1
Baseline, eyes closed	1 minutes	1
Open/close left or right fist	2 minutes	3
Imagine opening and closing left or right fist	2 minutes	3
Open/close both fists and or both feet	2 minutes	3
Imagine opening/closing both fists or both feet	2 minutes	3
Total runs:		14

B. Feature Extraction

In this study, four features including band power (BP), approximate entropy (ApEn), statistical features and wavelet-based features are extracted from the EEG data.

1) Band power: This is one of the most commonly used feature in the EEG studies. Firstly, the signal is decomposed into distinct frequency bands using the Fourier transform in (1). After that, power spectral density (PSD) is determined by taking the square of the magnitude of the Fourier transform using (2). Lastly, band power of EEG band is calculated by averaging the PSD over the frequency components.

$$\widehat{x_t}(\omega) = \frac{1}{\sqrt{T}} \int_0^T x(t)\, e^{-i\omega t} dt \quad (1)$$

$$S_x(\omega) = \lim_{T\to\infty} E[|\hat{x}_T(\omega)|^2] \quad (2)$$

Where, T represents period of signal, $S_x(\omega)$ represent PSD in and $|\hat{x}_T(\omega)|$ is the magnitude of the Fourier transform.

2) Approximate Entropy (ApEn): This is a time-domain feature that indicates the predictability and regularity of a signal [13]. Higher ApEn value indicates that there are more irregularities in the EEG signal. The steps for calculating ApEn are given below [14];

(a) The EEG signal is represented as a 1D array $[x_1, x_2, \ldots, x_n]$. (3) is applied to the array, and a vector y_i is obtained. For an EEG signal having length of n samples, total of $n - m + 1$ vectors are obtained.

$$y_i = [x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+m-1}] \quad (3)$$

Where, $i = 1, 2, 3, \ldots, n - m + 1$, m is the number of samples used for prediction.

(b) The maximum distance between 2 vectors, y_i and y_j is determined.

$$d[y_i, y_j] = max|y_{i+k} - y_{j+k}| \quad (4)$$

Where $k = 0, 1, 3, \ldots, m - 1$.

(c) The correlation integral, $C_i^m(r)$ is calculated for every vector using (5). It indicates the probability of the scalar of y_j over r of y_i.

$$C_i^m(r) = \frac{\{No.\text{ of } y_j \text{ that satisfy } d[y_i, y_j] < r\}}{(n-m)} \quad (5)$$

Where, r is the parameter of tolerance for comparison.

(d) Calculate the natural logarithms of $C_i^m(r)$ and take the average.

$$\varphi^m(r) = (n - m + 1)^{-1} \sum_{i=1}^{n-m+1} ln(c_i^m(r)) \quad (6)$$

(e) Repeat steps (i) to (iv) with $m = m + 1$, and compute $ApEn$ using equation (7).

$$ApEn\,(m, r) = \varphi^m\,(r) - \varphi^{m+1}(r) \quad (7)$$

3) Statistical features: Statistical features are also time domain features and include different quantitative statistics of the data. EEG signal is represented as a sequence of N sample points $X(n) = [x_1, x_2, \ldots, x_N]$. Six different statistical features are computed including mean, standard deviation, Interquartile Range, Median Absolute Deviation, skewness, and kurtosis as in (8) to (13) [9][15].

(a) Mean

$$\mu_x = \frac{1}{N}\sum_{n=1}^{N} X(n) \tag{8}$$

(b) Standard deviation

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X(n) - \mu_X)^2} \tag{9}$$

(c) Interquartile Range (IQR)

$$IQR = Q3 - Q1 \tag{10}$$

(d) Mean Absolute Deviation (MAD)

$$MAD = Median(|x_i - Median(X)|) \tag{11}$$

(e) Skewness:

$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^3}{\sigma_x^3} \tag{12}$$

Where, μ_x is the mean of the data.

(f) Kurtosis:

$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^4}{\sigma_x^4} \tag{13}$$

Where, σ_x is the standard deviation of the data.

4) Wavelet-based Feature: Wavelet transform is a method for extracting time-frequency features. A wavelet transform is a linear transformation that is represented by scaled and shifted by "mother wavelet". It decomposes the signal into a set of functions known as wavelet coefficients. Equation (14) is used to compute Wavelet-based features.

$$wt(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}(\,x(t)\,\phi^{*}\frac{t-\tau}{a})\;dt \tag{14}$$

Where, $wt(a,\tau)$ is the wavelet coefficients, $x(t)$ is the original signal, $\phi*\,(a,\tau)$ is the mother wavelet, a is the scale/ dilation parameter and τ is the shift/ translation parameter.

The wavelet features, consists of 2 frequency bands with 3 features each, which are mean, standard deviation, and energy.

C. Classifiers

Three different classification methods including KNN, SVM, and CNN are evaluated in the current study.

1) K-Nearest Neighbors (KNN): KNN is one of the simplest classification methods based on the principle that features will form clusters in feature space [16]. The features that are close together are the neighbors, and the major neighbors of a class is determined to specify the output label. 'K' in KNN is a parameter that refers to the number of nearest neighbors included in the voting process. To choose the best value for this parameter, cross-validation technique is used and the value with best performance on the validation set is selected.

2) Support vector machine (SVM): SVM is one of the commonly used classifiers for MI-EEG tasks. In general, SVM separates n-dimensional feature set into several planes that represent different output labels. SVM chooses the best hyper plane to distinguish the tasks performed by the user. The distance between each data point and the hyper plane is determined. Hyper plane is adjusted to maximize the distance with any data point.

3) Convolutional neural network (CNN): CNN is a special type of deep learning neural network model that is specialized to explore spatial information [15]. In BCI research, CNN has the advantage to capture the dependency relationship amongst the EEG signals from different electrodes. CNN model consists of convolutional layer, pooling layer, and fully connected layer.

III. RESULTS AND DISCUSSION

A. Performance Evaluation

For KNN and SVM classifiers, K-fold cross validation (with k=10) was used to evaluate the accuracy of the models. The dataset is divided into 10 parts and each time 9 parts were used for training and the last one for testing. The testing results from all iterations are averaged and recorded as the accuracy of the model. For CNN model, the training only went through one iteration. The dataset was divided into training data (80%) and testing data (20%). Training and testing accuracies were both recorded to observe the training condition. However, only testing accuracies were used for plotting the results.

B. Classification using different features separately

For every classifier, ApEn, BP, statistical and wavelet features were used as input for the classification separately.

The classification results using individual feature type are shown in Fig. 1. As we can see from Fig. 1, statistical features performed the best among all the feature types. The best classification was obtained using CNN model with statistical features resulting in 94.28% accuracy. SVM resulted in 94.05% and KNN 93.87% accuracy with statistical features. The worst classification was obtained using BP features and CNN classifier with accuracy of 50.13%. The lowest accuracy for SVM (58.41%) and KNN (61.54%) were also obtained using BP features.

CNN achieved highest classification accuracy with statistical features whereas the highest accuracies for the rest of features type were obtained with KNN. The CNN still has room for improvement by having more hyperparameter options. Moreover, the recorded result of CNN was limited due to the training duration where only 30 epochs were chosen. Including more epochs in the training process could potentially improve the results. Other than inherent information of the features, CNN also recognized the relationship of one electrode with others, hence able to learn more difficult patterns. Although SVM achieved lowest average accuracy, but it still obtained second best result with the accuracy of 94.05% with statistical features.

Statistical features were comprehended as features with most information by having highest average classification accuracy across different classifiers. The 2nd best feature, wavelet features, consisting of 2 frequency bands with 3 features each, which are mean, standard deviation, and energy. The single feature apEn resulted in a reasonable average accuracy of 86.54%. This indicates that randomness in the EEG data can discern imagining left and right-hand movements. In the current study the whole alpha EEG band (8-12 Hz) was chosen to calculate BP. We believe that power in smaller frequency ranges would increase the classification performance and hence the results with BP features.

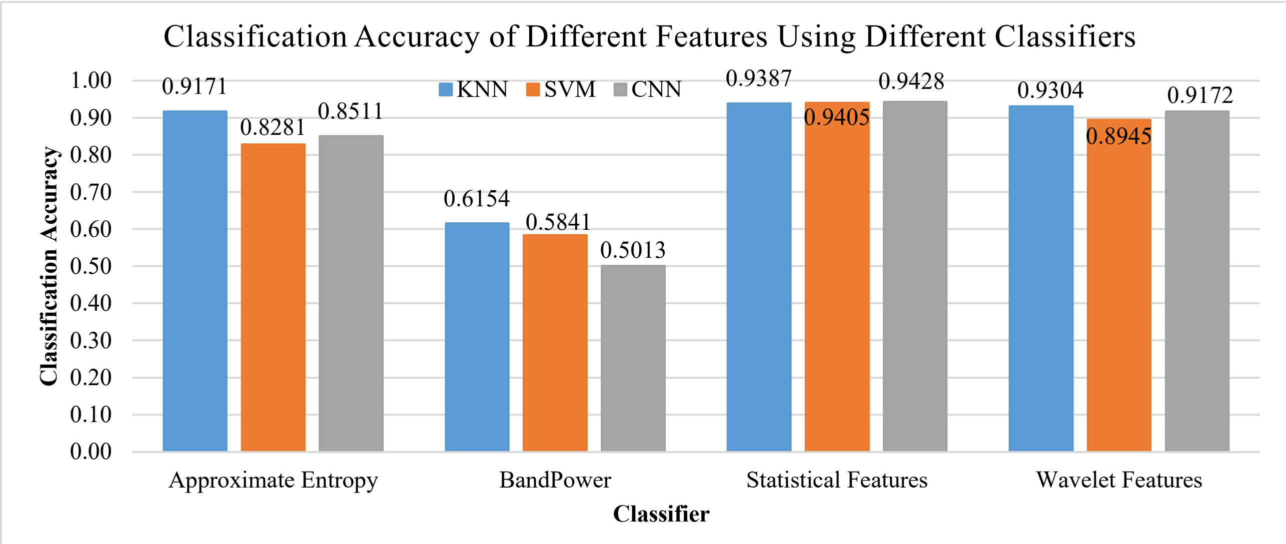

Fig. 1. Classification accuracy of different features using different classifiers for 64 electrodes with 0.8 overlapping ratio

C. *Classification result of combination of features*

The effect of using different features in combination is investigated and the results are shown in Fig. 2. Compared to single feature, overall classification accuracies were increased by combining the features. Highest accuracy was obtained by combining approximate entropy and statistical features and SVM as the classifier, with a result of 95.77%. Lowest classification was obtained by using approximate entropy and band power and KNN as the classifier, with a result of 81.61%. As we can see from the Fig. 2 not all the combinations increased the accuracy. Some models' accuracy dropped after including the band power into the feature set. For example, KNN model with approximate entropy as input had achieved 91.71% of accuracy at first but dropped significantly to 81.61% after combining with band power feature. This signifies the impact of unrelated features on the classification. The classification performance increased by combining the single features into grouped features except for the group that included band power. In general, it indicates that including more features provides more information to the classification system and hence increases the performance of the system. But this should be done cautiously because unrelated features might have bad impact to the classification process. In addition, having more features will also increase the computational cost of the system which can result in a negative impact for real-time BCI application. Fig 3 shows the training duration of different classifiers. It is observed that CNN and KNN have comparable speed whereas SVM takes much higher time for some features and hence is not suitable for real world cases.

D. *Classification using CNN without feature extraction*

In the previous section, classification was based on the features extracted before feeding into the classifiers. In this way we only extract certain information. But there is a chance of losing some other information which might be advantageous. To overcome this issue, CNN has the capability of identifying the features directly from the EEG data. In this section we report the results by directly feeding the EEG data into the CNN model. We evaluate different CNN model architectures with different number of EEG channels selected from the data.

Five different CNN models were built to classify the EEG data into left- and right-hand imagery tasks. Difference between the models were mainly number of convolutional filter and the kernel size. For model 1-3, the convolutional layers were 1-dimensional, while models 4 and 5 were 2-dimensional [17]. Number of epochs used for training was 10, and the batch size was 200. The number of channels was also reduced for classification as shown in Table 2. The classification accuracies are shown in Fig. 4. Highest testing accuracy was obtained using 3rd model with all electrodes used, which was 82.45%. lowest accuracy was 71.16% which was obtained using 2nd model with 2 electrodes used. The average accuracy across all 64 electrodes model was 80.82%, which was lower than single feature average classification accuracy of 82.18% with same 64 electrodes. However, the performance of CNN with raw input was better when the number of electrodes used were 6 and 2. Mean accuracy for single feature classifiers was 65.30% while raw input CNN resulted into average accuracy of 77.20% with 6 electrodes used. For 2 electrodes, raw input CNN achieved the average accuracy of 73.60%, whereas single feature classifiers average accuracy was 57.37%. Window sliding was not used in raw input CNN models whereas 80% overlap was used in the pre-computed feature classification. It indicates that CNN has the advantage to learn and extract useful features by itself with proper hyperparameters given. Avoiding the pre-computing feature extraction stage in CNN model is therefore highly advantageous and benefits the real-time BCI applications systems.

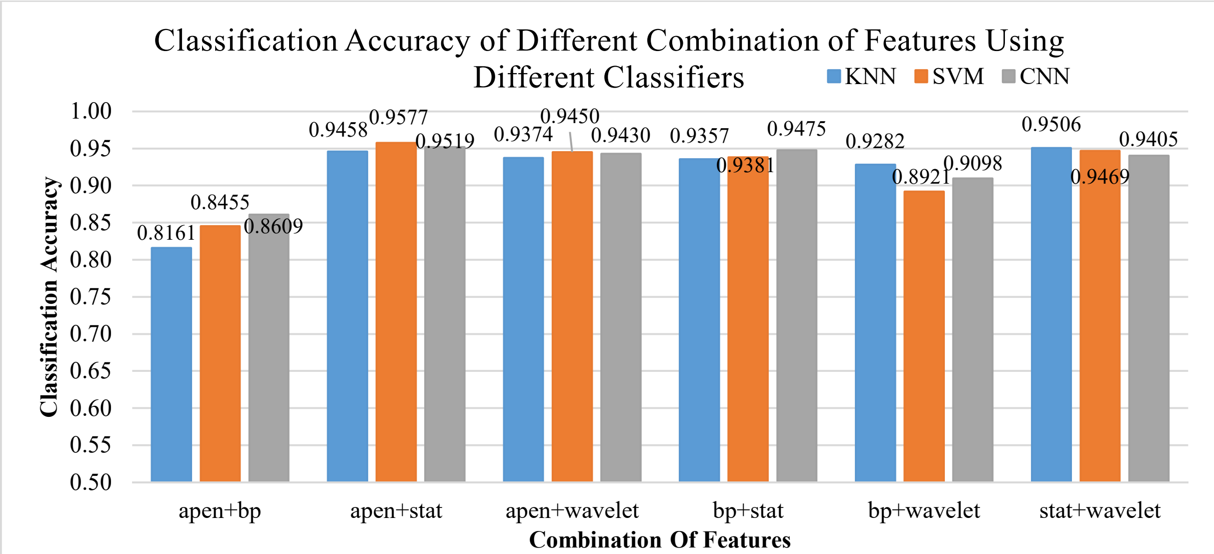

Fig. 2. Classification accuracy of different combination of features using different classifiers for 64 electrodes with 0.8 overlapping ratio

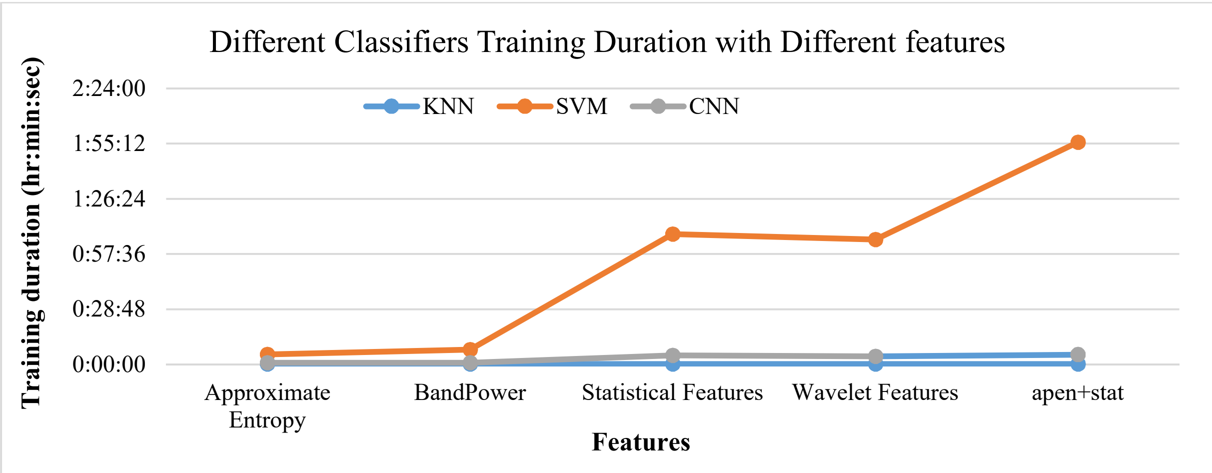

Fig. 3. Training Duration of Different Classifiers.

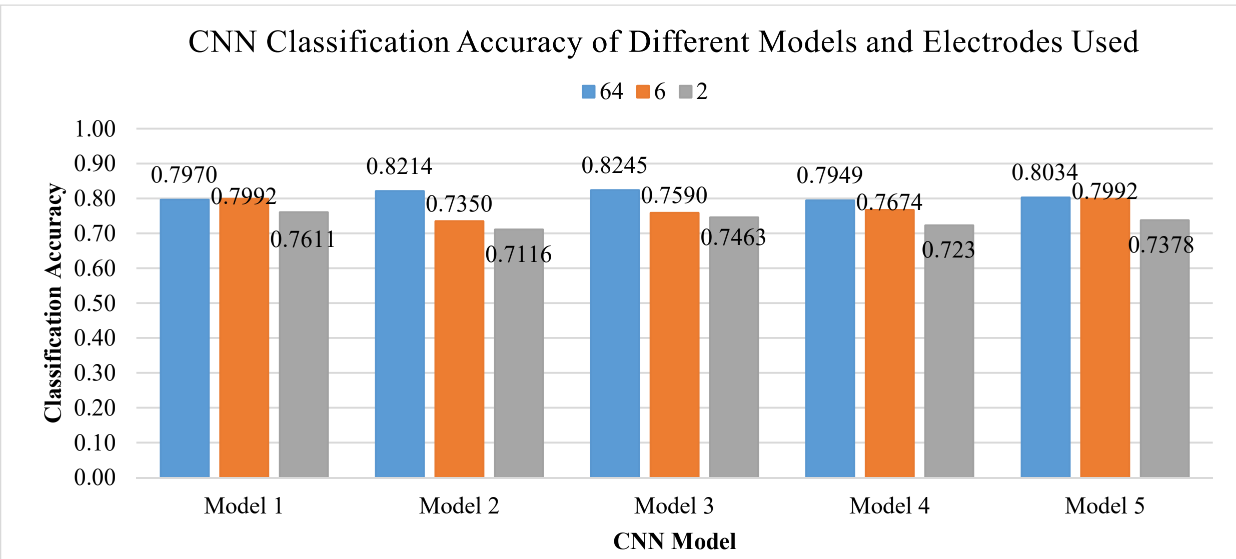

Fig. 4. CNN classification accuracy of different models and electrodes used

Table 2 Channels used for classification

Number of Channels	Channel Name
64	All
6	FC5-6, P7-8, AF3-4
2	C3, C4

IV. CONCLUSION

Publicly available EEG dataset was used to build MI classification model for BCI applications. We found that the EEG time-domain statistical features best classified the data. Though the combination of features increased the information to the classifiers, it is not always wise to include all the features for classification. Sometimes the inclusion of extra features just increases the computational cost of the model which could result in a negative impact for the real-time applications in terms of the system speed. It is concluded that using raw EEG as an input to the CNN model can eliminate the issue of pre-computing features with reasonable classification results.

ACKNOWLEDGMENT

The Universiti Tunku Abdul Rahman Research Fund (UTARRF) (Grant No. IPSR/RMC/UTARRF/2021-C2/H03), UTAR, Malaysia, has provided financial support for this research.

REFERENCES

[1] M. A. Lebedev and M. A. L. Nicolelis, "Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation," *Physiol. Rev.*, vol. 97, no. 2, pp. 767–837, 2017.

[2] Nisar, H., Khow, H.W. and Yeap, K.H., 2018. Brain-computer interface: controlling a robotic arm using facial expressions. Turkish Journal of Electrical Engineering & Computer Sciences, 26(2), pp.707-720.

[3] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[4] N. E. M. Isa, A. Amir, M. Z. Ilyas, and M. S. Razalli, "Motor imagery classification in Brain computer interface (BCI) based on EEG signal by using machine learning technique," *Bull. Electr. Eng. Informatics*, vol. 8, no. 1, pp. 269–275, 2019.

[5] M. Athif and H. Ren, "WaveCSP: a robust motor imagery classifier for consumer EEG devices," *Australas. Phys. \& Eng. Sci. Med.*, vol. 42, no. 1, pp. 159–168, 2019.

[6] D. Mzurikwao *et al.*, "A channel selection approach based on convolutional neural network for multi-channel EEG motor imagery decoding," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2019, pp. 195–202.

[7] N. Shajil, S. Mohan, P. Srinivasan, J. Arivudaiyanambi, and A. Arasappan Murrugesan, "Multiclass classification of spatially filtered motor imagery EEG signals using convolutional neural network for BCI based applications," *J. Med. Biol. Eng.*, vol. 40, no. 5, pp. 663–672, 2020.

[8] Nisar, H., Boon, K.W., Ho, Y.K. and Khang, T.S., 2022, June. Brain-Computer Interface: Feature Extraction and Classification of Motor Imagery-Based Cognitive Tasks. In 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (pp. 42-47). IEEE.

[9] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 910–926, 2020.

[10] Nisar, H., Hoe, T.C. and Nawaz, R., 2020, December. Reducing Sensors in Mental Imagery Based Cognitive Task for Brain Computer Interface. In 2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS) (pp. 1-10). IEEE.

[11] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N. and Wolpaw, J.R., 2004. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Transactions on biomedical engineering, 51(6), pp.1034-1043.

[12] Mahmood, D., H. Nisar, and Y.V. Voon. Removal of Physiological Artifacts from Electroencephalogram Signals: A Review and Case Study. in 2021 IEEE 9th Conference on Systems, Process and Control (ICSPC 2021). 2021. IEEE.

[13] R. Nawaz, H. Nisar, and V. V. Yap, "The effect of music on human brain; Frequency domain and time series analysis using electroencephalogram," *IEEE Access*, vol. 6, pp. 45191–45205, 2018.

[14] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 110–117, 1995.

[15] N. Sinha, D. Babu, et al., "Statistical feature analysis for EEG baseline classification: Eyes Open vs Eyes Closed," in *2016 IEEE region 10 conference (TENCON)*, 2016, pp. 2466–2469.

[16] M. Rashid *et al.*, "Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review," *Front. Neurorobot.*, p. 25, 2020.

[17] Cheah, K.H., Nisar, H., Yap, V.V. and Lee, C.Y., 2020. Convolutional neural networks for classification of music-listening EEG: comparing 1D convolutional kernels with 2D kernels and cerebral laterality of musical influence. Neural Computing and Applications, 32(13), pp.8867-8891.