Online multiclass EEG feature extraction and recognition using modified convolutional neural network method

Haider Abdulkarim[1], Mohammed Z. Al-Faiz[2]
[1]Department of Electrical and Communication Engineering, Uruk University, Iraq
[2]Al Nahrain University, Iraq

Article Info

ABSTRACT

Many techniques have been introduced to improve both brain-computer interface (BCI) steps: feature extraction and classification. One of the emerging trends in this field is the implementation of deep learning algorithms. There is a limited number of studies that investigated the application of deep learning techniques in electroencephalography (EEG) feature extraction and classification. This work is intended to apply deep learning for both stages: feature extraction and classification. This paper proposes a modified convolutional neural network (CNN) feature extractor-classifier algorithm to recognize four different EEG motor imagery (MI). In addition, a four-class linear discriminant analysis (LDR) classifier model was built and compared to the proposed CNN model. The paper showed very good results with 92.8% accuracy for one EEG four-class MI set and 85.7% for another set. The results showed that the proposed CNN model outperforms multi-class linear discriminant analysis with an accuracy increase of 28.6% and 17.9% for both MI sets, respectively. Moreover, it has been shown that majority voting for five repetitions introduced an accuracy advantage of 15% and 17.2% for both EEG sets, compared with single trials. This confirms that increasing the number of trials for the same MI gesture improves the recognition accuracy.

Corresponding Author:

Haider Abdulkarim
Department of Electrical and Communication Engineering
Uruk University, Baghdad, Iraq
Email: haiderasadi@daad-alumni.de

1. INTRODUCTION

Brain computer interface (BCI) is an emerging field, in which the signals generated by brain during certain tasks, are recorded and used to control specific systems, such as moving a wheeled chair or an artificial arm. BCI system has three main stages: Signal acquisition, feature extraction and classification. The accuracy of the BCI system mainly relies on the performance of the classification technique used. Many proposed methods have been introduced in this field. Basically, there are three main feature extraction techniques: time features, frequency features and statistical features [1].

The classification step, on the other hand, uses the extracted features to classify the electroencephalography (EEG) signals into the intended BCI actions. Linear discriminant analysis (LDA), support vector machine (SVM) and neural networks (NN) are examples of the most common classifiers used in BCI. In 1986, the term deep learning (DL) was introduced by Dechter [2]. Her work focused on constraint recording whenever dead-ends arise as a process of learning. Then the recorded information is used to avoid future conflicts. She used DL term to describe the extraction of all the possible information out of a dead-end. It basically means recording all the explored search space in order to improve the learning performance.

DL has been implemented in many applications to extract high level features from stochastic data, such as speech and image. One of its most used applications is convolutional neural networks (CNN). This class of deep learning uses convolution filters as the first feature extraction layer. CNN is widely used in image processing and classification algorithms. However, CNN application in motor imagery (MI) has not yet been intensively investigated. CNN is one class of deep-learning neural networks, where there is more than one hidden layer. CNN is used to extract features that are hidden in the data, which usually cannot be extracted by shallow neural networks. In BCI applications, EEG signals related to MI movements are captured and then their features are extracted. Afterwards, the desired action that matches the target class is recognized and the target output is performed (e.g., moving a wheeled chair to a target point according to the MI intension of a disabled person). CNN can be adopted to be used in both EEG feature extraction and MI classification.

Fedjaev [3] investigated two approaches: long short-term memory (LSTM) and CNN to classify two MI classes (right and left hand). The proposed LSTM configuration consisted of single layer and 128 LSTM memory cells. The suggested CNN architecture has four layers, 1024-time samples, window size of 4 seconds and three EEG channels. The proposed LSTM model scored 84.2% while the proposed deep CNN scored 92.2%. Dose et al. [4] implemented a temporal and spatial convolutional layer for feature extraction and fully connected layer for classification. The proposed model consisted of two one-dimensional convolutional layers, with each layer consists of 40 filter kernels. The first layer implement convolution while the second layer applies a spatial filter along the EEG channels. The layer output is weighed linear combinations of the single channel readings. Therefore, the dimensionality of the data along the EEG channel dimension is reduced to one. Afterwards, mean pooling is applied to further reduce data size. The output is finally passed to a fully connected layer for classification. An accuracy of 87.0% was reached for 2-class but the accuracy degraded for three and four class, with 76.6% and 65.7%, respectively.

Yang et al. [5] proposed the use of augmented common spatial pattern (ACSP) on BCI competition IV EEG data set IIa, reaching maximum accuracy of 69.27% for four-class discrimination. Jingwei et al. [6] implemented CNN with multi-scale feature selection to learn high level features (deep motor features). They reached 100% accuracy using "Project BCI - EEG motor activity data set" with reduced training time of 4.125 seconds. However, the proposed model requires very large training set of 100 trials to classify 360 trials test data, which is not always feasible in real-time BCI application demands. Lee and Choi [7] implemented a one-dimensional convolutional filter to capture time-frequency features. They proposed a three-layer model to classify two classes of BCI competition, gaining a mean accuracy of 78.9%. Zhang et al. [8] proposed seven-layer DL model with orthogonal array experiment technique to optimize hyper parameters like the number of layers and nodes. They reached a mean accuracy of 93%. In addition, they incorporated the model into a simulated robot model for smart home applications. Tang et al. [9] proposed five-layer CNN model to extract feature and classify EEG feature of two class MI (left vs right hand) with accuracy of 86.4%. Shahin et al. [10] used Deep Neural Network to discriminate between patients with insomnia, using only one and two EEG channels, reaching 86% and 92% accuracy, respectively. Nurse et al. [11] used the Truenorth chip supplied with CNN algorithm to decode two classes MI movements (left vs. right), with 81% accuracy. Zhang et al. [12] proposed a combined gait-EEG authentication system. They exploited the fact that each person has a unique EEG signature in addition to the uniqueness of the human walking activity and they achieved an identification accuracy of 98.3%. Wang et al. [13] applied CNN feature extraction on BCI competition III dataset IVa (3 classes) and used a weak classifier on the extracted features. They gained an average accuracy enhancement of 5% with respect to Neural Network and LDA. Sturm et al. [14] implemented layer-wise relevance propagation (LRP) which shows individual classification decisions of a DNN. LRP decomposes the output in terms of input samples. They implemented the proposed method on two class (left vs right hand) with accuracy close to CSP-LDA. Chiarelli et al. [15] combined functional near infrared spectroscopy (fNIR) with DNN to increase the accuracy of binary classification, attaining increased performance compared to using DNN only.

Heyden [16] compared between the analysis of variance (ANOVA), genetic algorithm techniques as feature extractor, and linear vs radial basis support vector machine (SVM) as a classifier, to classify two MI EEG signals. The best accuracy was found for ANOVA, with 67.7%. Zeng et al. [17] compared between CNN and EEG-Conv-R, which combines CNN with recent deep residual learning. The proposed CNN architecture composed of eight layers. The task was to classify two mental states of a car driver; attentive vs distracted. The average accuracy for CNN and EEG-Conv-R achieved 82.9% and 84.3%, respectively. Stober et al. [18] implemented recursive neural network (RNN) binary model to classify 2D from 3D objects, shown to subject. The maximum accuracy of 72% is reached. Dai et al. [19] used a combined representation of time, frequency and channel information of EEG signals and a combination of CNN for feature extraction and variational auto encoder (VAE) for classification. However, only two classes are considered.

Nagabushanam et al. [20] used long-short-term memory (LSTM) to reach a classification accuracy of 78.9. LSTM is based on recurrent neural network (RNN) which has a low computation overhead.

However, the proposed model was a binary classifier. Djamal *et al.* [21] implemented combined wavelet transformation and CNN to identify a stroke vs. no stroke occurrence (a binary classifier) with 90% accuracy. Ali *et al.* [22] used deep learning to diagnose autism spectrum disorder, reaching 80% accuracy. Djamal *et al.* [23] used combined wavelet transformation and RNN. The wavelet transform is used to extract time-frequency features and the RNN is used as the classifier. The goal was to classify four direction movement of a drone besides the focus vs. non focus status. The obtained accuracy was 79.6%.

Schirrmeister *et al.* [24] made an intensive study of deep convolutional neural networks with different structures, and showed that the deep learning classification performed as good as Filter Banks common spatial pattern (FBCSP) in terms of accuracy (84% vs 82.1%). Tabar and Halici [25] implemented a combined system of CNN and a stacked autoencoder and then used time, frequency and location information to classify a two-class EEG dataset of the BCI competition IV. The dataset was obtained by capturing Electrocorticography (ECoG) signals during the flexion of five fingers and reached accuracy of 90% because ECoG has higher Signal to noise ratio than EEG. Asim *et al.* [26] used a combined machine learning feature extraction and SVM classifier to detect Alzheimer (binary classification) and achieved 94%. Ajaj *et al.* [27] implemented short time fourier transform (STFT) to reduce the size of data entering the neural network. They reached 99.8% accuracy using weighted KNN approach. However, the tested approach was a two-class problem (focus vs non focus). Niroshana *et al.* [28] used fine-grained segment in multi scaled entropy along with CNN to determine the sleeping status of the subject. The classification problem was to identify the sleeping stage. The proposed method reached an average accuracy of 92.2%. Doborjeh *et al.* [29] used spiking neural networks (SNN) to classify the attention of customers to a stimuli, with 89.9% as a binary classification problem. Hajinoroozi *et al.* [30] used a novel channel-wise CNN to detect the drivers attention status, reaching 84.4% accuracy. Hussein *et al.* [31] used LSTM along with fully connected NN to detect the occurrence of epileptic seizers, reaching 100% accuracy. Ieracitano *et al.* [32] converted the power spectral densities (PSD) of the 19-channel EEG signals into a gray scale image and then applied CNN method to classify the severity stage of Dementia. They reached 89.8% accuracy in binary classification and 83.3% accuracy in a three-degree classification.

Deep learning techniques aim to extract high level features from a large time samples size. Those features cannot be extracted directly from an input matrix with high number of samples. In addition, the classification task mainly relies on the accuracy of those features. Usually, deep learning is used in image classification and speech recognition. The CNN is one major branch of deep learning techniques, which has been widely used in image processing. However, it has only been recently introduced to the field of BCI using EEG. Therefore, this paper proposes a modified CNN model to extract the EEG features and then feed these features to a neuro-classifier to classify unknown multi-class EEG gestures.

This paper introduces a modified CNN feature extraction and classification algorithm able to classify 4 different EEG gestures. This work used online captured EEG as the data source. This guarantees that the proposed model will integrate efficiently with BCI applications, especially high demanding real time applications. This paper is organized as follows. In section 2.1, the proposed modified CNN feature extractor structure is explained. Then, in section 2.2, the shallow neural network classifier structure is described. For the purpose of performance comparison with EEG classification algorithms that solve more than two classes, the multi class LDA is revised in section 2.3. In section 3, the EEG session recording process is described. In addition, the CNN classification results are discussed and compared to multi-class LDA algorithm.

2. RESEARCH METHOD
2.1. Proposed CNN model

As mentioned in section I, the DL has been used in image processing and classification. In brief, the CNN is composed of the following layers [19]:
− Convolution layer(s)
− Pooling layer(s)
− Shallow neural network layer(s)

The convolutional layer (CL) uses a kernel filter, which is convolved with the original array elements using the dot product operation. In computer vision, the data elements represent the pixels of the image, which is a two-dimensional array (for each Red, Green, and Blue channels), while the kernel filter is 3 by 3 or 5 by 5 [32]. The function of the 2-D kernel filter is to detect both the edges as well as the points of changes in the image domain. This is done by applying different Kernel filter types, such as averaging, root mean square (RMS) and sharpen filter. Each of these Kernel filters has one goal: to emphasize the differences in adjacent pixels and Hence to extract a reduced-size set of features for an image. However, in EEG, the structure of data is different: multi-channel readings are composed of T trials by M rows × N columns [33], where T is the number of EEG trials (for the same EEG gesture), M is the number of channels and N is the

time samples. Since M << N, the 2-D filters that are used in the field of image processing are not applicable to extract the features of EEG signals. In this proposed CNN model, the convolution filter is basically a 1-D mean filter, as in (1):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \tag{1}$$

where x_i is the ith time sample and n is the total number of samples for which the mean is calculated. As shown in (1) is applied to a window of size W_C. The filter is then shifted by a step size (*stp*). Therefore, the final size of the convolved EEG input is given by:

$$M \times \{\frac{(N-W_C)}{stp} + 1\} \tag{2}$$

where W_C is the window size (in samples). In this paper, W_C is set to 128 samples. For a sampling rate of 128, this window is equivalent to 1 second. This time window is found (through experimental trials) to be sufficient to capture the EEG activity feature for a single channel. The output size of this layer is:

5 *trials*×14 *channels*×129 *averaged time samples*

The second layer, Pooling layer_1, extract the RMS value of the elements of each row within a window size $W_P = 64$, according to (3):

$$X_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_{ci}^2} \tag{3}$$

where x_{ci} is the ith element in the output of the convolution layer, n is the selected window size, which is 64. The size of samples is reduced by a factor of 64 and the output size of the Pooling layer can be formulated as:

$$M \times (N_2 - W_p) + 1 \tag{4}$$

where N_2 is the column size of the convolutional layer and W_p is the pooling layer window size. Therefore, the output size of Pooling_layer_1 is 5×14×2. In the next layer, Pooling layer 2, the mean value of a window size of each two samples is calculated. The final feature matrix size for each class is 5×14, which is converted to 1-D vector of 70 samples. The complete structure of the deep learning layers is shown in Figure 1. The output of the Pooling layer_2 is fed to a shallow neural network as a training set. Each set is a vector of 70 element features and the corresponding output labels are (0, 1, 2 and 3) for Class$_0$, Class$_1$, Class$_2$ and Class$_3$, respectively. The NN layers architecture is described in the next section.
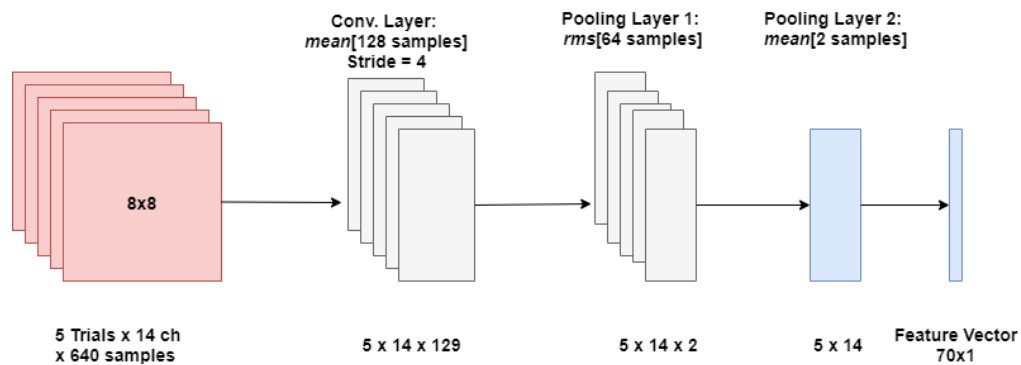

Figure 1. The proposed deep learning feature extraction structure for each class

2.2. Neural network classifier

After features are extracted in section II, the size of time features has suitable structure to be fed to the next stage: Neuro-Classifier. The proposed NN structure is:
- 70 input neurons

− 150 hidden neurons
− 1 output neuron

The hidden size might be reduced after weight updates, according to the eliminated weights. The nodes of the hidden layer are fully connected to both the input layer and to the output layer nodes, respectively, with random initial weights. Because the desired output is either 0 or 1, each neuron has a unipolar activation function, log sigmoid, defined as:

$$f(X) = logsig(w^T X) \tag{5}$$

Training phase: The neural network weights update is based on back propagation (BP) with momentum term algorithm, where the weights are updated based on the error rate in the previous iteration (Epoch). The weights of the connections of the nodes of each layer are calculated as [34]:

$$\Delta W_n = \eta \sigma y^T + \mu W_{n-1} \tag{6}$$

where η is the learning rate, σ is the error signal vector of a layer and y^T is the transposed input vector to that node. μ is the rate at which the previous weight (W_{n-1}) contributes to the current weight (W_n). To choose the best values for η and μ, the learning phase is repeated for different values and the convergence time is observed. The learning error rate is calculated using the mean square error between the desired output and the NN classifier output. Then, upon evaluating the learning rate error, the following values are chosen based on best error performance and shortest convergence time:

η=0.3, μ=0.6 and error_goal=10^{-2}

In the training phase, NN weights are updated at the end of each epoch until the error reaches the minimum stopping condition or a maximum of 1000 epochs are occurred. Afterwards, the trained neural network weights are saved and used in the next stage, classification. In this stage, unknown-class time features (56 elements) extracted from the deep learning stage is applied to the trained NN to get the desired classification result.

Recognition phase: After the NN is trained with the training set, the final weights are stored so that no training is needed for the new inputs (unknown class). Then, the unknown-class data set is fed as an input to the first layer (convolutional layer) and the same process is repeated (feature extracted using deep learning). The final output is a feature vector of size 70×1. Each trial feature vector belongs to a predefined class. Therefore, the trails are fed to the trained NN (within a loop) and the output of each loop represents the classification result, rounded to either 0 or 1. Since the NN is a binary classifier, a binary-tree classification strategy is proposed because the recognition task has more than two classes. The binary-tree classifier is depicted in Figure 2.
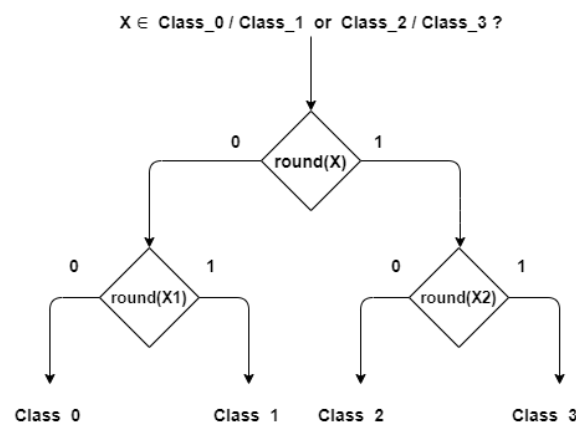

Figure 2. Four-class binary-tree classifier

At the first level, feature vectors of $Class_0$ and $Class_1$ are stacked in a single feature matrix. Another feature matrix is composed for $Class_2$ and $Class_3$. Then, the trained binary classifier (Level_1) decides to

which combination the test data belongs (based on rounded 0/1 output). Then, based on Level_1 output, one of two binary classifiers is chosen: $Class_0$ vs $Class_1$ or $Class_2$ vs $Class_3$. The binary classifier will decide to which class the test features belong. The Class decision is formed according to Table 1.

Table 1. Binary tree classifier for a four-class problem

Class0 _Class1 (0) vs Class2 _Class3 (1)	Level 1 Output	Binary Classifier Output	Final Class Decision
0	Class0 _vs_Class1	0	Class0
0	Class0 _vs_Class1	1	Class1
1	Class2 _vs_Class3	0	Class2
1	Class2 _vs_Class3	1	Class3

2.3. Multi class LDA classifier

For the purpose of comparison with the proposed CNN model, the binary LDA is revised and then multi-class LDA classifier model is derived according to [34]. In classification problems, the samples are normally non-linearly separable. Therefore, a projection on other space is needed. The LDA calculates the projection vector w according to the generalized Rayleigh quotient [35]:

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \qquad (7)$$

where S_B is the between-class covariance array and S_W is the within-class covariance array. To solve for a solution of w that maximizes $J(w)$, Fisher Linear Discriminant is used:

$$w = S_W^{-1}(m_1 - m_0) \qquad (8)$$

where m_1 and m_0 are the mean for $class_1$ and $class_0$, respectively. The binary LDA used with two-class EEG shows good accuracy of 90% [36]. The LDA can be extended to be used in m-class problems, where the $class_k$ is selected based on the minimum Euclidean distance:

$$class_k = arg_k \min(euk\ dist(xw^T, \mu_k w^T)) \qquad (9)$$

where x is the unknown class feature vector, w^T is the projection vector and μ_k is the mean of $class_k$.

3. RESULTS AND DISCUSSION

EMOTIV EPOC+headset is used to collect EEG raw EEG signals. The headset supports 14 channels, each with a sampling rate of 128 samples/sec [37]. The raw data is recorded using MATLAB 2019B. The output data from the headset is already filtered with a built-in digital 5th order sinc bandpass-filtered (0.16~43) Hz, with two digital notch filters at 50 Hz and 60 Hz. Each of the proposed classifiers (CNN and LDA) operates in two phases:

The training phase: The test subject is asked to rest his arms and all body for 15 minutes before recording session begins. A cue on the screen is shown to the subject to indicate the starting of record. Each EEG trial record time is 5 seconds. During the trial capturing period, the subject will perform an EEG gesture according to $Class_0$ (Rest). A stopping cue is shown to indicate the ending of trial recording. The subject then repeats the same class gesture for the remaining four trials using the same procedure. A total of five trials for the same class is recorded. These trials are used as a training data for $Class_0$, which is fed to the classifying algorithm. The same recording procedure applies to the remaining three classes: $Class_1$, $Class_2$ and $Class_3$. The recognition phase: same as training phase, except that the subject picks one of the four gestures on a random basis, for three repetitions only. The recorded data is fed to the classifying algorithm as an unknown-class gesture. The EEG training and recognition trials recording procedure are summarized in Table 2.

Table 2. Class-vs-gesture trials

Class	Corresponding Gesture	Training trial repeat (each 5 sec)	Recognition trial repeat (5 sec)
Class0	Rest	5	
Class1	Blinking	5	
Class2	Left eye	5	1,3 and 5 repeats (Each of four classes)
Class3	Right eye	5	

To calculate the classification accuracy, the subject selects a set of random actions. A total of 100 trials are tested. The performance accuracy of the classification algorithm is calculated as:

$$Classification_{Accuracy} = \frac{(1-False\ classif)}{True_{classif}-False\_classif} \qquad (10)$$

Figure 3 shows the proposed CNN model classification accuracy, calculated for each set of EEG gestures: Arm rest, both eyes blinking, Left eye blink and Right eye blink. Each class has 4 trial sets. Each set is classified while the other three sets are used as training sets. Within the same set, the classification accuracy is calculated in three options: single, three and five trials. The class / trials with the highest accuracy are shown. Finally, the average accuracy is calculated among all the four classes with the best set accuracy is chosen. This accuracy is calculated for single, majority 3 and majority 5 repetitions, respectively. The same description applies to Figure 4, using multi-class LDA.

A comparison between Figures 3 and 4 shows that CNN algorithm outperforms the LDA for the same set of EEG gestures with a 27.1% enhancement in average accuracy. To validate the performance of the proposed CNN model, another MI motion set is chosen (arm rest, eye brows up and down, blinking both eyes and both hands opening and closing). It is shown that CNN still performs better than multi-class LDA with a margin of 15.55%, as shown in Figures 5 and 6, respectively. This complies with the fact that LDA is best suited for binary classification problems which have two classes, while CNN can be efficiently adapted and learned to multi-class BCI applications.
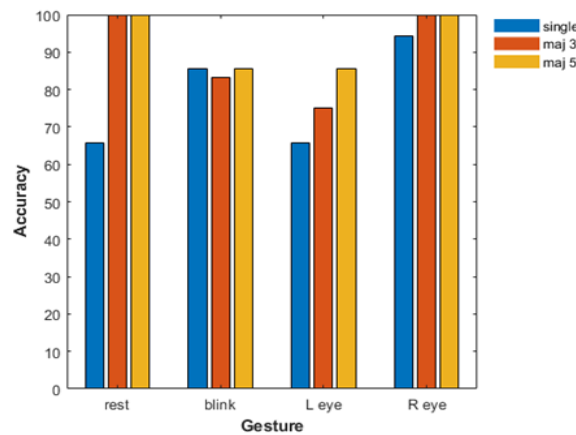

Figure 3. CNN classification accuracy (4 classes). EEG set: rest, blinking, left eye blink, right eye
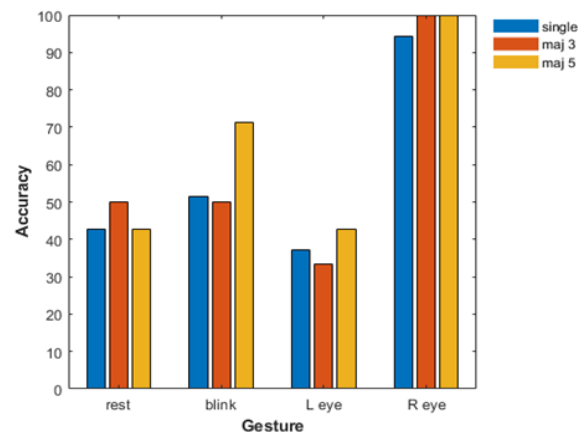

Figure 4. LDA classification accuracy (4 classes). EEG set: rest, blinking, left eye blink, right eye blink
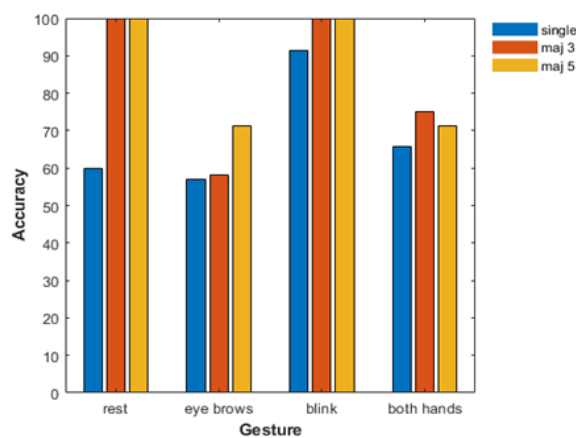

Figure 5. CNN classification accuracy (4 classes). EEG set: rest, eyebrow, blink, and both hands
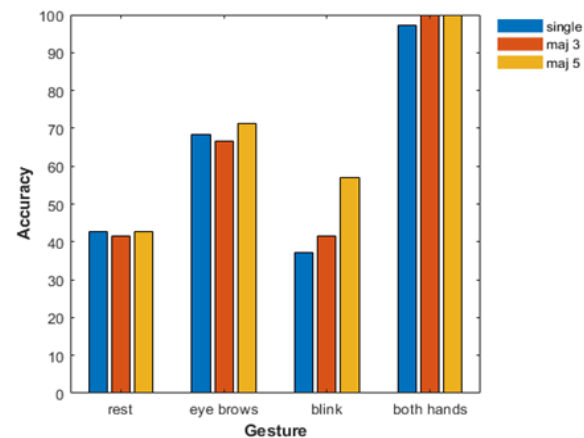

Figure 6. LDA Classification Accuracy (4 classes). EEG set: rest, eyebrow, blink, and both hands

In addition, increasing the number of trial repetitions for the same gesture significantly increase the classification accuracy. For example, in Figure 3, there is an average accuracy increase of 15% when five gesture repetitions are used against single repetition. The majority voting [20], [38] is used to decide for the class with the most frequent occurrence in each 5 consecutive trials. Therefore, there is a very good accuracy advantage when using repeated gestures combined with majority voting. For example, for the following output vector:

$$\{Class_0, Class_2, Class_1, Class_2, Class_3\}$$

The winning output class is 2, accroding to majority voting rule. When comparing the individual class accuracies of Figures 3 and 4, the right eye EEG gestures has the highest accuracy compared to other gestures. This is caused by the fact that the test subject EEG signal related to the right eye is stronger than other EEG related to other gestures. This leads to a better feature vector and consequently, more accurate classification. In addition, in Figure 5, both eyes blinking has the highest average accuracy. The average computation time is calculated using the MATLAB command (CPU time), which calculates the total CPU time elapsed by the algorithm code. The average computation time for the CNN classifier is 7.9 seconds, compared to 0.35 seconds for LDA. Although the LDA classifier has less computation time, the CNN classifier provides better accuracy for the same EEG data set. Table 3 shows that the proposed algorithm outperforms other solutions proposed by the most related works in the field, for which the same number of classes used and CNN technique is implemented.

Table 3. Related work accuracy comparison

Work	EEG channels	Method	Avg. Accuracy
Dose *et al.* [4]	64	CNN	65.7%
Djamal *et al.* [23]	4	LSTM based on RNN	79.6%
This work	14	CNN	89.2%

Table 4 compares the average accuracies of both the proposed CNN model and multi-class LDA, using two MI sets. It shows that CNN performance is better than LDA for the same set and the same number of trials. This table also indicates that the performance using MI set_1 is better than set_2.

Table 4. CNN vs. LDA performance comparison

Trials	Accuracy		
	Single	Majoriy 3	Majority 5
Set 1 (CNN)	77.8	89.5	92.8
Set 1 (LDA)	56.3	58.3	64.2
Set 2 (CNN)	68.5	83.3	85.7
Set 2 (LDA)	61.3	62.4	67.8

4.   CONCLUSION

In this paper, a multi-class EEG convolutional neural network classification algorithm for motor imagery is proposed. This classifier is adopted from image classification into the EEG recognition field. Major modifications are done so that the classifier can fit the stochastic nature of EEG signals. This work has used EMOTIV EPOC headset to capture online EEG data, instead of using online available EEG datasets. This is an important feature that makes this model satisfy online BCI applications, such as moving a disabled wheeled chair or manipulating an artificial arm. In addition, this model achieved small computation time, which is an advantage in real-time applications. It could be further reduced using more powerful on-board processing units. This algorithm is capable of labeling four different classes with very good average accuracy of 92.8% for one EEG set and 85.7% for another set. This result outperforms many similar proposed DL techniques, in terms of accuracy and number of classes. In addition, a four-class LDA classifier model is implemented based on extension of binary LDA and the accuracy is compared to the proposed. The results show that the CNN classifier outperforms the LDA with a margin of 27.1% and 15.5% for set 1 and set 2, respectively. In addition, it has been proven that majority voting combined with repeated trails significantly improves the classification accuracy. The improvement is 15% and 17.2% for the 2 sets, when using majority-5 voting against single trial.

In addition, is has been found that certain EEG MI gestures provide more individual classification accuracy than others. For example, right eye blinking (for set 1) and both eyes blinking (set 2) gives higher

recognition rate than other gestures. This outcome corresponds with the test subject EEG nature and might differ from other subjects. Therefore, a future investigation on more subjects is required to determine the general selection rule for the best EEG gesture. Although the CNN classifier model computation time is larger than that of LDA, CNN classifier model has better classification accuracy for the same EEG set, which makes it better suited for high accurate BCI applications demands. The importance of a multi-class EEG classifier over binary classifier is that it allows the implementation of variant tasks of BCI applications. For example, the algorithm presented in this work can drive a robot into four directions: forward, backward, left, and right.

The proposed algorithm, like other BCI-EEG classification methods, suffers from performance degradation when the number of classes increases. This can be solved by using combined time-frequency features of selected EEG channels (channels with maximum entropy values could be used). Those features are then fed to the CNN algorithm described earlier with minor modifications. Another limitation of the algorithm is that the training EEG data for each class must be recorded and stored in advance. This must be done only once for each subject. If a new person wants to use the same EEG-MI system, the training operation must be repeated. However, the algorithm still capable of classifying new EEG data for the same subject as long as the same user is using the trained BCI system. As a future improvement to the current work, common spatial pattern (CSP) can be used to select the channels with maximum signal to noise ratio. This could enhance the overall accuracy and allows to increase the number of recognized classes. In addition, channel mutual information could be used as another criteria for channel selection. Furthermore, combined time-frequency features can be extracted and use the same proposed CNN structure with minor modifications. Wavelet transform can also be used in this context. Then the channels with the highest mutual information can be selected and neglect other channels that has noise-like contribution.

REFERENCES

[1] F. Lotte *et al.,* "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018, doi: 10.1088/1741-2552/aab2f2.

[2] R. Dechter, "Learning While Searching In Constraint-Satisfaction-Problems," *AAAI'86: Proceedings of the Fifth AAAI National Conference on Artificial Intelligence*, 1986, pp. 178-183.

[3] J. Fedjaev, "Decoding EEG Brain Signals using Recurrent Neural Networks," Technische Universität München (Neurowissenschaftliche Systemtheorie), 2017.

[4] H. Dose, J. S. Møller, S. Puthusserypady, and H. K. Iversen, "A deep learning MI-EEG classification model for BCIS," *2018 26th European Signal Processing Conference (EUSIPCO)*, Rome, Italy, 2018, pp. 1676-1679, doi: 10.23919/EUSIPCO.2018.8553332.

[5] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Italy, 2015, pp. 2620-2623, doi: 10.1109/EMBC.2015.7318929.

[6] J. Liu, Y. Cheng, and W. Zhang, "Deep learning EEG response representation for brain computer interface," *2015 34th Chinese Control Conference (CCC)*, Hangzhou, China, 2015, pp. 3518-3523, doi: 10.1109/ChiCC.2015.7260182.

[7] H. K. Lee and Y. S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequecy image," *2018 International Conference on Information Networking (ICOIN),* Chiang Mai, Thailand, 2018, pp. 906-909, doi: 10.1109/ICOIN.2018.8343254.

[8] X. Zhang, L. Yao, C. Huang, Q. Z. Sheng, and X. Wang, "Intent recognition in smart living through deep recurrent neural networks," *International Conference on Neural Information Processing-ICONIP 2017,* vol. 10635, pp. 748-758, 2017.

[9] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11-18, 2017, doi: 10.1016/j.ijleo.2016.10.117.

[10] M. Shahin, B. Ahmed, S. T. Ben Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, "Deep Learning and Insomnia: Assisting Clinicians with Their Diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1546-1553, 2017, doi: 10.1109/JBHI.2017.2650199.

[11] E. Nurse, B. S. Mashford, A. J. Yepes, I. Kiral-Kornek, S. Harrer, and D. R. Freestone, "Decoding EEG and LFP signals using deep learning: Heading truenorth," *CF '16: Proceedings of the ACM International Conference on Computing Frontiers*, 2016, pp. 259-266, doi: 10.1145/2903150.2903159.

[12] X. Zhang, L. Yao, C. Huang, T. Gu, Z. Yang, and Y. Liu, "DeepKey: An EEG and Gait Based Dual-Authentication System," arXiv:1706.01606 vol. 9, no. 4, pp. 1-22, 2017, [Online]. Available: http://arxiv.org/abs/1706.01606.

[13] Q. Wang, Y. Hu, and H. Chen, "Multi-channel EEG classification based on fast convolutional feature extraction," *International Symposium on Neural Networks-ISNN 2017*, vol. 10262, pp. 533-540, 2017, doi: 10.1007/978-3-319-59081-3_62.

[14] I. Sturm, S. Lapuschkin, W. Samek, and K. R. Müller, "Interpretable deep neural networks for single-trial EEG classification," *Journal of Neuroscience Methods*, vol. 274, pp. 141-145, 2016, doi: 10.1016/j.jneumeth.2016.10.008.

[15] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid EEG-fNIRS brain-computer interface: Application to motor imagery classification," *Journal of Neural Engineering*, vol. 15, no. 3, 2018, Art. no. 036028, doi: 10.1088/1741-2552/aaaf82.

[16] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cognitive Neurodynamics volume*, vol. 12, no. 6, pp. 597-606, 2018, doi: 10.1007/s11571-018-9496-y.

[17] A. S. Greaves, "Classification of EEG with Recurrent Neural Networks," *Arxiv*, vol. 18, no. 1989, pp. 1-23, 2015, [Online]. Available: http://arxiv.org/abs/1511.04306.

[18] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors,* vol. 19, no. 3, pp. 551-567, 2019, doi: 10.3390/s19030551.

[19] P. Nagabushanam, S. Thomas George, and S. Radha, "EEG signal classification using LSTM and improved neural network algorithms," *Soft Computing*, vol. 24, no. 13, pp. 9981-10003, 2020, doi: 10.1007/s00500-019-04515-0.

[20] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, and M. K. M. F. Alif, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, pp. 91-99, 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[21] E. C. Djamal and R. D. Putra, "Brain-computer interface of focus and motor imagery using wavelet and recurrent neural networks," *TELKOMNIKA (Telecommunication Computing, Electronics and Control)*, vol. 18, no. 5, pp. 2748-2756, 2020, doi: 10.12928/TELKOMNIKA.v18i5.14899.

[22] R. T. Schirrmeister *et al.,* "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391-5420, 2017, doi: 10.1002/hbm.23730.

[23] Y. R. Tabar, and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, 2017, Art. no. 016003, doi: 10.1088/1741-2560/14/1/016003.

[24] Y. Asim, B. Raza, A. K. Malik, S. Rathore, L. Hussain, and M. A. M. Iftikhar, "A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning," *International Journal of Imaging Systems and Technology*, vol. 28, no. 2, pp. 113-123, 2018, doi: 10.1002/ima.22263.

[25] S. M. Isuru Niroshana, X. Zhu, Y. Chen, and W. Chen, "Automatic sleep stage classification based on convolutional neural networks," *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech),* Osaka, Japan, 2019, pp. 275-276, doi: 10.1109/LifeTech.2019.8883961.

[26] Z. Gholami Doborjeh, M. G. Doborjeh, and N. Kasabov, "Attentional Bias Pattern Recognition in Spiking Neural Networks from Spatio-Temporal EEG Data," *Cognitive Computation*, vol. 10, no. 1, pp. 35-48, 2018, doi: 10.1007/s12559-017-9517-x.

[27] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Processing: Image Communication*, vol. 47, pp. 549-555, 2016, doi: 10.1016/j.image.2016.05.018.

[28] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25-37, 2019, doi: 10.1016/j.clinph.2018.10.010.

[29] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96-107, 2019, doi: 10.1016/j.neucom.2018.09.071.

[30] J. Jin *et al.,* "Bispectrum-Based Channel Selection for Motor Imagery Based Brain-Computer Interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 10, pp. 2153-2163, 2020, doi: 10.1109/TNSRE.2020.3020975.

[31] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal Feature Selection Method of CSP Based on L1-Norm and Dempster-Shafer Theory," *IEEE Transactions on Neural Networks and Learning Systems,* pp. 1, 2020, doi: 10.1109/tnnls.2020.3015505.

[32] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *Journal of Neural Engineering*, vol. 16, no. 3, 2019, Art. no. 031001, doi: 10.1088/1741-2552/ab0ab5.

[33] M. Simonovsky, and N. Komodakis, "Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 29-38, doi: 10.1109/CVPR.2017.11.

[34] K. S. Narendra, and K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4-27, 1990, doi: 10.1109/72.80202.

[35] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: An experimental investigation," *Knowledge and Information Systems*, vol. 10, no. 4, pp. 453-472, 2006, doi: 10.1007/s10115-006-0013-y.

[36] M. Z. Al-Faiz and H. Abdulkarim, "Offline Linear Discriminant Analysis Classfication Of Two Class Eeg Signals," *Iraqi Journal of Information & Communications Technology (IJICT),* vol. 2, no. 3, pp. 1-10, 2019, doi: 10.31987/ijict.2.3.71.

[37] Emotiv, "brain computer interface technology Emotiv EPOC User Manual," 2016. [Online]. Available: https://www.192-168-1-1-ip.co/manuals/13886.pdf.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," p. 738, 2000, [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471056693.html.

BIOGRAPHIES OF AUTHORS

Haider Abdulkarim was born in Baghdad, Iraq, 1981. He received his B.Sc. and M.Sc. in Communications and Signal Processing, from Baghdad University, Iraq and Ilmenau University of Technology, Germany, in 1999 and 2012, respectively. Since 2013 he has been an assistant lecturer of Digital Communications theory. He is currently a PhD candidate at the departmet of Information and Communication, college of Information Engineering, Al Nahrain University, Iraq and his research interests includes EEG signal processing and classification, Deep Learning and Convolutional Neural Networks.

Mohammed Al-Faiz received his M.Sc. in Automation of Command, Electronic Computers and PhD in Technical Cybernetics from Brno Military Academy, Czech Republic, in 1983 and 1987, respectively. Since 1992, he has been the Dean and head of many scientific departments and also a senior member of the IEEE. He has many contributions and researches in Robotic and Medical Engineering fields and has supervised 19 PhD students and more than 70 M.Sc. and B.Sc. students. He has more than 50 published research articles and two written academic books. He is currently an emeritus professor at the college of Information Engineering, Al-Nahrain University, Iraq.