City, University of London Institutional Repository

Classification of Motor Imagery EEG Signals Based on Deep Autoencoder and Convolutional Neural Network Approach

JAMAL F. HWAIDI AND THOMAS M. CHEN

Department of Electrical and Electronic Engineering, City, University of London, London EC1V 0HB, U.K.

Corresponding author: Jamal F. Hwaidi (jamal.hwidi@city.ac.uk)

ABSTRACT The technology of the brain-computer interface (BCI) employs electroencephalogram (EEG) signals to establish direct interaction between the human body and its surroundings with promising applications in medical rehabilitative services and cognitive science. Deep learning approaches, particularly the detection and analysis of motor imagery signals using convolutional neural network (CNN) frameworks have produced outstanding results in the BCI system in recent years. The complex process of data representation, on the other hand, limits practical applications, and the end-to-end approach reduces the accuracy of recognition. Moreover, since noise and other signal sources can interfere with brain electrical capacitance, EEG classifiers are difficult to improve and have limited generalisation ability. To address these issues, this paper proposes a new approach for EEG motor imagery signal classification by using a variational autoencoder to remove noise from the signals, followed by a combination of deep autoencoder (DAE) and a CNN architecture to classify EEG motor imagery signals which is capable of training a deep neural network to replicate its input to output using encoding and decoding operations. Experimental results show that the proposed approach for motor imagery EEG signal classification is feasible and that it outperforms current CNN-based approaches and several traditional machine learning approaches.

INDEX TERMS Electroencephalography, deep autoencoder, convolutional neural network, variational autoencoder, motor imagery.

I. INTRODUCTION

Brain-computer interface (BCI), also known as the brain-machine interface (BMI), creates a non-muscle channel that allows the human body to communicate directly with external devices [1], [2]. There are three types of BCI paradigms: invasive, partially invasive, and non-invasive. The most powerful procedures, such as visual or motor implants [3], [4], are the most invasive, but they come with many risks associated with surgery, such as scar tissue and infections. On the other hand, non-invasive procedures such as electroencephalography (EEG) [5] can help with medical diagnosis and research while also addressing real-world issues. Electrocorticography (ECoG) [6] is a semi-invasive method that only requires surgery to implant devices on the brain's surface.

The most extensively studied non-invasive BCI is EEG, which is relatively inexpensive and simple to carry out, while

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

providing fine temporal resolution. As a result, it is a common method for analyzing and monitoring changes in brain electrical activity.

The frequency range of brain oscillations is typically between 0.5 and 40 Hz. Based on these measurements, EEG signals are divided into five rhythms: Delta δ (0.5-4 Hz), Theta θ (4-7 Hz), Alpha α (8-13 Hz), Beta β (14-30 Hz), and Gamma waves γ (>30 Hz) that are all present in different parts of the brain [7]. EEG aids in the detection of a variety of brain abnormalities such as sleeping disorders, emotional variance, and seizure detection, and researchers have recently proposed user identification systems [8], neuromarketing [9] and rating prediction systems [10].

EEG-based BCI has been studied for its effects on various paradigms, including exogenous and endogenous characteristics. One of the most common modes is motor imagery (MI). While some oscillating activity in the brain's sensorimotor cortex corresponds to specific imagination, MI refers to subjects performing an action by imagining a specific part of

their body (such as the left and right hands, as well as the feet) instead of moving it [11].

Machine learning technology is frequently used to categorize and identify these MI tasks. The MI-BCI system can be used to control external devices such as wheelchair controls or neural prostheses for people with disabilities, as well as to assist healthy people with difficult tasks, such as controlling devices, automatic driving, and epilepsy diagnosis. Because of the low cost, high availability, and lack of any manual support with EEG signals, motor imagery BCI systems have become increasingly powerful intelligent machines.

The MI-BCI system has traditionally been divided into five phases: acquisition of signal data, preprocessing of data, feature extraction, classification, and device control interface [12]. MI-EEG signal collection, signal digitalisation, and data storage are all part of the data acquisition phase. Filtering, cleaning, and transformation of data are all part of the data preprocessing phase. Discriminative features are extracted from EEG signal data that contain useful information during the feature extraction phase. The extracted features are used as input to train machine learning models in the classification phase. Different signals and MI tasks can be classified using the trained models. Finally, in the device control interface phase, the categorized signals are converted into commands to control devices like robots and home appliances [13]–[15].

In the field of MI-EEG research, a wide range of feature extraction and classification techniques have recently been used. The Common Spatial Patterns (CSP) and its variants, such as the filter-bank CSP (FBCSP), have progressed considerably in the classification of MI. Deep learning techniques, on the other hand, do not rely on handcrafted features.

In this paper, a novel deep learning architecture for MI EEG signal classification based on DAE and CNN is presented. The CNN convolution kernel is trained using an unsupervised deep autoencoder. The main benefit of combining a deep autoencoder with a CNN is that the DAE is a nonlinear feature extraction technique that is used before classification of a high-dimensional dataset to remove redundant information that exists in the EEG signal. It also has a significant advantage over other competitors because it is capable of learning a deep neural network that is trained to replicate its input to its output through encoding and decoding operations.

The CNN convolution kernel is trained using an unsupervised deep autoencoder. The deep autoencoder is a fully connected three-layer neural network with an unsupervised training method because a label is not required during training. Encoder training aims to learn better representations of input data in order to train better classifiers.

The remainder of the paper is organised as follows: Section II provides a comprehensive review of the main approaches developed for MI-EEG. Section lll describes the proposed DAE and CNN approach. The experimental results and discussion of the approach are presented in Section IV, followed by conclusions in Section V.

II. RELATED WORK

A variety of machine learning algorithms and feature extraction methods have been studied in the field of MI-EEG in order to overcome the limitations of small datasets and low signal-to-noise ratios. The research background can be formed into two groups of traditional machine learning approaches and modern deep learning-based ones. Following is a more in-depth discussion of these methods.

A. TRADITIONAL MACHINE LEARNING APPROACHES

Many techniques for MI-EEG classification have been proposed, with the traditional method of hand-crafted feature extraction in the time and frequency domains being one of them. One of the most popular and powerful methods for feature extraction is the Common Spatial Patterns (CSP) [16], [17], which has been used in a variety of other approaches, for example Filter-Bank CSP (FBCSP) [18], [19], Sub-band CSP (SBCSP) [20], sparse CSP [21], and discriminative filter bank CSP (DFBCSP) [22] that has achieved good accuracy on public datasets such as BCI Competition IV datasets 2a and 2b.

Many researchers use time frequency signal processing methods for feature extraction, such as the short time Fourier transform (STFT) which was used by Tabar and Halici [23] to extract information about time, frequency, and location from raw EEG signals while also converting them to images, continuous wavelet transform (CWT) which was used by Lee and Choi [24] to convert EEG signals into spectrums of time and frequency, and empirical mode decomposition (EMD) [25]. Furthermore, principal component analysis (PCA) and independent component analysis (ICA) are examples of dimension reduction techniques that is used to improve MI task recognition performance [26]–[28].

In addition, traditional classifiers such as Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) have been used to discriminate hand-crafted feature vectors. A transfer function for dimensionality reduction was proposed by Ji and Ye [29] as a consolidated framework for generalised Linear Discriminant Analysis (LDA). The framework describes the characteristics of various algorithms as well as their relationships.

The SVM can also produce promising results when combined with the right features [30]. Kumar *et al.* [31] proposed a mutual information based frequency band selection bases on CSP features and LDA is used to further reduce dimensionality before SVM is used to classify the patterns.

Park and Chung [32] developed a method for extracting more discriminative CSP features for an SVM classifier using method of channel optimal selection, resulting in 88.62% accuracy. Furthermore, Kumar *et al.* [33] published a follow-up study in which, before classifying with SVM, they use a Long Short-Term Memory (LSTM) network for temporal filtering and LDA for spatial filtering.

Islam *et al.* [34] proposed a multiband tangent space mapping with sub-band selection (MTSMS) approach where the tangent features of a multichannel electroencephalogram (EEG) signal are estimated on each sub-band after it is

decomposed into multiple sub bands. Sub-bands with features capable of improving motor imagery classification accuracy are selected using a mutual information analysis-based effective algorithm. Feature space is created by combining the obtained features of selected sub-bands and to reduce the features dimension, a principal component analysis-based approach is used, the data is then classified using SVM.

Despite the fact that these traditional approaches have been successful in recognising MI, designing accurate BCIs remains a challenge, leaving room for further research in this area.

B. DEEP LEARNING APPROACHES

In the field of EEG signals, the traditional machine learning approach is commonly used. However, its EEG signal processing performance and accuracy are insufficient. To address this issue, researchers have begun to investigate the feasibility of employing many deep learning approaches in the analysis of EEG signals.

The CNN was used by Li *et al.* [35] to classify MI signals. They began by extracting primary features such as EEG channel dependency as well as temporal features. After that, CNN is used to extract high-level features. Moreover, based on deep CNN, Chaudhary *et al.* [36] developed a method for recognising MI tasks in a BCI system. The DCNN is specifically used to classify EEG signals from right hand and foot MI-tasks. The proposed method uses time-frequency (T-F) approaches to convert the input EEG signals into images. Short-time Fourier Transform (STFT) and continuous-wavelet-transform (CWT) are two time-frequency approaches that are commonly used. The DCNN stage is used to apply the images of MI-tasks EEG signals after T-F transformation. AlexNet, a DCNN model that has been pre-trained, is investigated for classification. Lee and Choi [37] constructed 2D images are used to train a CNN model by using CWT.

Zhang *et al.* [38] developed a novel approach by combining deep learning and data augmentation. To improve classification accuracy for the small training dataset, empirical mode decomposition method on EEG frames, two neural networks were used to generate new artificial frames, train the weights, and classify two classes of motor imagery signals: a CNN and a wavelet neural network (WNN) which is a new type of neural network that replaces convolutional layers with wavelets employed. For MI classification, Wu *et al.* [39] proposed the paralle (MSFBCNN) multiscale filter bank convolutional neural network. In a layered end-to-end network structure, a feature-extraction network extracts temporal and spatial features. On small datasets, you can use this method to train an individual model for inter-subject classification, a setup and fine-tuning strategy for networks is used to improve the transfer learning ability.

Amin *et al.* [40] proposed a multi-layer CNNs (MCNN) method for improving EEG MI classification accuracy by combining CNNs of various characteristics and architectures (CCNN) when capturing spatial and temporal features

from raw EEG data, different convolutional features are used. The MCNN and CCNN outperform all state-of-the-art machine learning and deep learning techniques.

Lu *et al.* [41] developed a new deep learning method based on restricted Boltzmann machines (RBM). For comparison, the frequency representation of EEG recordings was obtained using the Fast Fourier transform (FFT) and wavelet package decomposition (WPD) separately. The frequency domain data is then used to train a deep belief network (DBN) made up of RBM.

Xu *et al.* [42] employed transfer learning to train their CNN, which was designed to classify MI signals. By transferring the parameters of the VGG-16 (visual geometry group) pre-trained network to their proposed CNN model, they were able to take advantage of it.

Alazrai *et al.* [43] employed the Choi-Williams distribution (CWD) transformation on EEG signals to create 2D images for CNN input. The energy distribution in both the time and frequency domains is reflected in these images.

Zhao *et al.* [44] proposed a multi-branch CNN for MI signal classification. The 3D representation is created by converting EEG signals into a 2D array sequence that keeps the spatial distribution of sampling electrodes intact. For the 3D representation, the classification strategy and multi-branch 3D CNN are designed specifically. Three deep learning models for decoding movements in motor imagery precisely from raw EEG signals have been proposed by Tayeb *et al.* [45]: (1) a short-term memory that is both long and short (LSTM); (2) a spectrogram-based CNN; and (3) a recurrent convolutional neural network (RCNN). State-of-the-art machine learning techniques were outperformed by deep learning models in terms of classification performance, which could pave the way for the development of new robust EEG signal decoding. Using the CNN-based BCI, they successfully controlled a robotic arm in real time.

Kwon *et al.* [46] developed a large EEG database based on motor imagery (MI) and proposed a deep CNN-based framework that is subject-independent. A large-scale MI database is used to represent the overall brain signal patterns by using spectral–spatial input generation which shows superior performance of the feature representation combined with a method based on deep neural networks.

III. METHODOLOGY

The proposed system framework is depicted in Fig 1. Each subject data was downloaded separately making a total of 10 data which is mixed randomly according to trial that is later divided into one piece of data for the test set and the remaining nine are for the training set. The test and training set combined data from several subjects. Each trial yielded the MI-EEG raw signals from nine pairs of symmetrical electrodes spread across the motor cortex, with each pair's signals forming a sample. Moreover, each extracted signal contains significant amount of noise. As a result, the noise was filtered using a variational auto encoder (VAE).
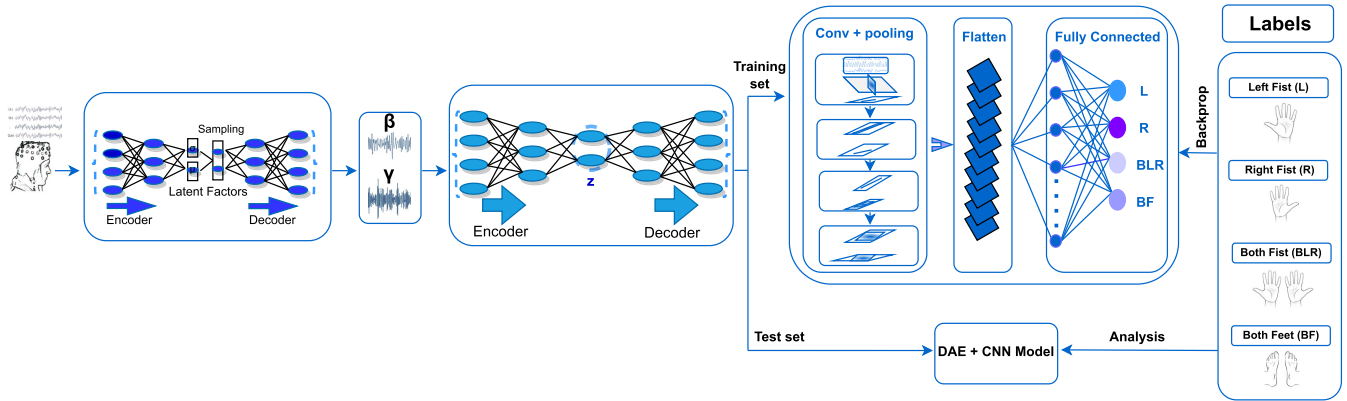
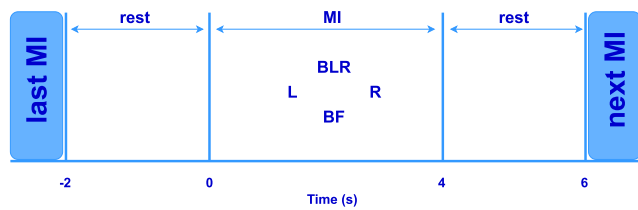**FIGURE 1.** The proposed system framework.

FIGURE 2. Timing diagram of the trial.

The proposed CNN model was trained in conjunction with a deep auto decoder, which aids in the representation of the input at the output layer so that both are as similar as possible. The EEG features were learned using a 5-layer CNN, and the dimensions were reduced using 4-layer max pooling. MI was divided into four categories by the FC layer: left fist, right fist, both fists, and both feet. The best training model can then be found by comparing it to the four types of labels. Finally, the test set was used to assess the model's validity.

A. DATASET DESCRIPTION

The Physionet dataset [47] used in this paper was recorded by the BCI2000 system developers [48]. It consists of over 1,500 EEGs (one and two minutes) recordings using a sampling rate of 160 Hz from 109 different subjects. For each subject, four MI tasks were completed: left fist (T1), right fist (T2), both fists (T3), and both feet (T4), with each MI task requiring 21 trials. Fig 2 illustrates the trial's timing diagram.

The trial begins at t = −2 seconds, and the subject relaxes for 2 seconds. The target appears on the screen at t = 0, and the symbols are described below:

1) L denotes the motor imagination of the left fist being opened and closed.
2) R denotes the motor imagination of the right fist being opened and closed.
3) BLR denotes the motor imagination of both fists being opened and closed.
4) BF denotes the motor imagination of both feet being opened and closed.

The MI task was given to the subject for four seconds. At t = 4s, the target vanishes, indicating that the trial is over. A new trial begins after a 2 second break [49]. The motor imagination runs for about 4 seconds each time with a sampling frequency of 160 Hz. As a result, each electrode's effective data size per test is 640. A pair of symmetric electrodes are present in the sample, and their data is connected in series. Therefore, the sample size is 1,280.

A total of 84 trials were completed by each subject on each MI task, totalling 21 trials per MI task. The dataset in this study was subjected to 10-fold cross validation. Each subject's trials were divided into ten parts. Each task class has its own set of requirements, two test trials were used and the remainder as training. As a result, the test set contains 8 trials, while the training set contains 76 trials. In 10 subjects' datasets (S1∼S10), there are 840 trials, 760 for training and 80 for testing. Furthermore, each trial yielded 9 samples. For the training of models and the validation of generalisation performance, 10 subject dataset with 7,560 samples were selected in this experiment.

B. PREPROCESSING

The data is subjected to a signal amplification and filtration process at the time of acquisition. To demonstrate how to segment the data stream, data segmentation was used. The following steps were used to preprocess the EEG dataset:

1) The data was sampled at a rate of 128 Hz.
2) The Variational Autoencoder (VAE) was used to remove EEG noise using a blind source separation technique, based on prior EEG noise reduction research [50].
3) The discrete wavelet transform (DWT) was used to split EEG signals into bands; beta and Gamma waves were considered in this experiment.
4) A common reference was used to average the data and the pre-trial segments of three seconds were removed due to the fact that the data was divided into 60 second trials.
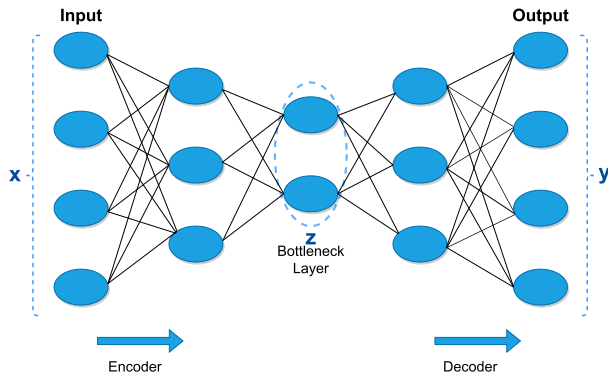
FIGURE 3. Deep Autoencoder.

C. DEEP AUTOENCODERS

As shown in Fig 3, an autoencoder is a fully connected artificial neural network system with a bottleneck layer [51]. In this paper, a deep autoencoder is used to extract ERP and morphological features from EEG signals in order to help capture neural activities related to both sensory and cognitive processes.

The encoder and the decoder are the two blocks that make up an autoencoder. At the bottleneck layer, the encoder's goal is to convert a higher-dimensional input feature vector into a lower-dimensional representation. At the decoder end of the autoencoder, the bottleneck features are transformed into a higher dimensional representation, and the input and output features drive the autoencoder learning, ensuring that the bottleneck layer presents a lower dimensional representation of the input features. The operation of encoding can be described as follows:

$$z = f(\theta; x) = l(Wx + n) \qquad (1)$$

where the input feature vector x is represented by the bottleneck feature vector z, which propagates through hidden layers. $\theta = \{W, n\}$. W is the weights of the network, n is the network biases, and l is a linear or non-linear activation function.

At the output stage, the bottleneck feature vector z, which propagates through hidden layers at the decoder, is mapped to the higher dimensional representation y as

$$y = j(\theta'; z) = l(W'z + n') \qquad (2)$$

where $\theta' = \{W', n'\}$. As a result, the DAE output can be expressed as a function of the encoder and decoder stages weights and biases, namely $\{\theta, \theta'\}$ and written as $y = j(\theta'; (f(\theta; x)))$. The DAE parameters θ and θ' are optimised to bring y as close to input/target x as possible while also maximising $p(x \mid y)$. Moreover, mean square error (MSE) backpropagation between target x and network output y is used to optimise the autoencoder parameters.

The deep autoencoder was chosen because it learns a compressed representation of the input in order to reconstruct it later, making it suitable for dimensionality reduction. It is

made up of an encoder and a decoder, which are used to deal with or mitigate the effects of the curse of dimensionality, and they feed the learned features to a standard CNN classifier. The features are derived from data pertaining to the stimulus onset of self-face images by averaging voltages in 20 non-overlapping time-windows with a width of 10 sample points, yielding a 5×20 matrix per trial, which only uses the selected channels and time intervals. The sparse method was used to expand the 5×20 matrixes to 20×20 matrixes, i.e., adding zero to the blank space, and a 20×20 feature map was generated for each trial to organise the features for CNN training.

D. CNN

The dataset consisted of 10 subjects and 7,560 samples, with 6,840 samples in the training set and 720 samples in the test set. The number of layers in the structure, as well as their parameters were determined through a series of experiments. Five layers of CNN and four layers of max pooling have been identified.

This paper used the deep convolution neural network as shown in Fig 4, which is useful for identifying key features vector.

Table 1 shows the CNN architecture that was chosen: The input layer is the first layer; the convolutional layers are the second, third, fifth, seventh, and ninth layers; the max pooling layers are the fourth, sixth, eighth, and tenth layers; and the fully connected layer is the eleventh layer.

The CNN input data format is $S \times E$, where S denotes the sampling amount for each channel and E denotes the number of electrodes used. In this paper, the output of the DAE were used as $S = 488$ and $E = 64$.

Down sampling is the core of the pooling process. Max pooling was chosen, which is achieved by taking the largest value of the neighbourhood's attributes. It can better extract feature information by suppressing the situation where network parameter error causes the predicted mean value to move. An algorithm [52] is shown in the max-pooling equation (3) below:

$$z_{x,w} = \max_{(i,j) \in l_{x,w}} y_{i,j} \qquad (3)$$

where $z_{x,w}$ is the result of the kth feature map's pooling operator and $y_{i,j}$ is a pooling region that represents a local neighbourhood around the position at (i, j) of $l_{x,w}(x, w)$.

The FC layer is deployed after feature extraction to improve the network's nonlinear mapping capacity. It interprets global data and combines local features learned from the convolutional layer to create global classification features. In the same layer, there are no connections between neurons in this layer, and every neuron in the preceding layer is linked to all neurons in the layer before it.. The formula is as follows:

$$y_j^{(l)} = a\left(\sum_{i=1}^{n} x_i^{(l-1)} \times w_{ji}^{(l)} + b^{(l)}\right) \qquad (4)$$

where n denotes the previous layer's number of neurons. In this layer, the weight of the connections between neurons

j and neurons I in the preceding layer is $w_{ji}^{(l)}$, j neurons have a bias of $b^{(l)}$, and the activation function is a.

A softmax layer with four neurons $[t^1, t^2, t^3, t^4]$ that represent the four categories generates the FC layer's output. It maps the outputs of several neurons to the interval (0, 1), which could be thought of as a possible outcome of multiclassification. The formula is as follows:

$$t^n = \arg\max \left(\frac{e^{t^n}}{\sum_{n=1}^{4} e^{t^n}} \right) \tag{5}$$

Rectified linear units (ReLu) [53] were chosen as the activation functions as the gradient vanishing problem during training is less likely and training convergence is faster as follows:

$$f(x) = \begin{cases} 0, & \text{if } x > 0 \\ x, & otherwise \end{cases} \tag{6}$$

Furthermore, the proposed CNN architecture uses the batch normalisation (BN) algorithm to improve classification accuracy. The batch normalisation method involves performing normalisation processing (mean value 0 and standard deviation 1) before each layer of the network input for each batch of data. That is, for any neuron in this layer, $\hat{x}^{(k)}$ uses the following formula (presuming the k-th dimension):

$$\hat{x}^{(k)} = \frac{x^{(k)} - E\left[x^{(k)}\right]}{\sqrt{Var\left[x^{(k)}\right]}} \tag{7}$$

where $x^{(k)}$ is the kth neuron's original input data, $E[x^{(k)}]$ is the mean of the kth neuron's input data p, and $\sqrt{Var\left[x^{(k)}\right]}$ is the standard deviation of the data in the kth neuron.

Batch normalisation adds more constraints to the data distribution, improving the model's generalisation ability. After normalisation, the mean and standard deviation of the input distribution are both forced to be 0 and 1. Reconstruction of the transformation and learnable parameters γ and β are introduced in a particular implementation to re-distribute the data as it was originally:

$$z^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \tag{8}$$

where $\gamma^{(k)}$ and $\beta^{(k)}$ represent the input data distribution's variance and deviation, respectively.

The formula for the batch normalised network layer's complete forward normalisation process is as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i' \tag{9}$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left(X_i' - \mu\right)^2 \tag{10}$$

$$X_i^{norm} = \frac{X_i' - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{11}$$

$$\widetilde{X}_i = \gamma X_i^{norm} + \beta \tag{12}$$

where X_i' is the input, μ is the mean, and the variance is σ^2.
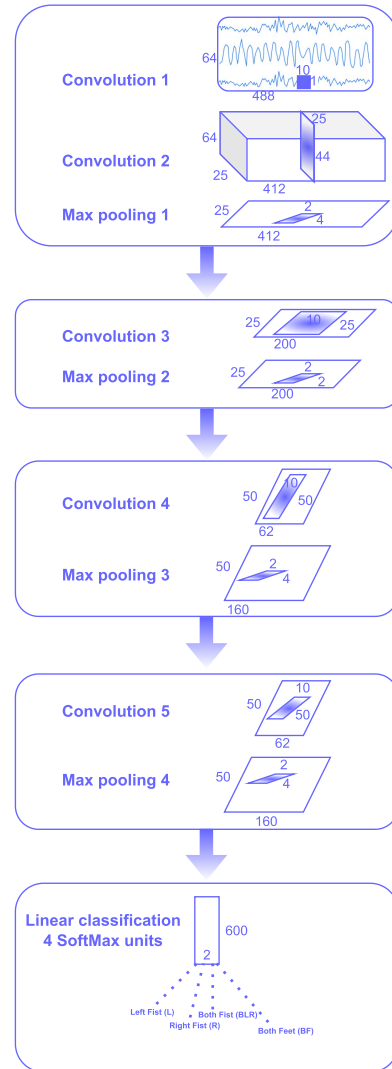

FIGURE 4. The proposed CNN architecture is depicted in this diagram. 5 convolutional layers, 4 max-pooling layers, and a fully connected layer make up the model. For this work, Adam was used as the optimization algorithm.

The training method used was mini-batch training, which divides the total number of the sample that were trained into reduced chunks and instead of learning one single sample, all parameters are updated after learning one mini-batch. The batch size used was 64.

The Adam optimisation algorithm was used to minimise the loss function using a constant of 1×10^{-5} as the learning rate of the network in this study.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, DAE was used to extract the features and the performance of the features evaluated by using the CNN classifiers was explored.

In the classification step, the dataset was subjected to 10-fold cross validation, with 90% of the data being used to train the CNN model, which was then tested for robustness

TABLE 1. Proposed CNN architecture.

Layer	Type	Maps	Size	Kernel Size	Stride	Padding	Activation	Parameters
Input	Input	1	488×64	-	-	-	-	-
C1	Convolution1	32	412×25	2×4	1	VALID	3072	208
A1	Activation	32	-	-	-	-	ReLU	-
C2	Convolution2	32	412×25	2×2	1	VALID	1568	416
B1	Batch Normalization	32	-	-	-	-	-	-
A2	Activation	32	-	-	-	-	ReLU	-
P1	Max pooling1	32	210×25	2×4	2	SAME	-	-
C3	Convolution3	32	200×50	2×2	1	VALID	1600	3216
A3	Activation	32	-	-	-	-	ReLU	-
P2	Max pooling2	64	160×50	2×2	2	SAME	-	-
C4	Convolution4	64	120×100	2×2	2	SAME	400	48120
B2	Batch Normalization	64	-	-	-	-	-	-
A4	Activation	64	-	-	-	-	ReLU	-
P3	Max pooling3	64	90×100	2×4	2	SAME		
C5	Convolution5	128	50×200	2×2	1	VALID	120	10164
B3	Batch Normalization	128	-	-	-	-	-	-
A5	Activation	128	-	-	-	-	ReLU	-
P4	Max pooling4	128	6×200	2×2	2	SAME	-	-
Flatten	Flatten	-	600	-	-	-	-	-
F1	Fully Connected	-	200	-	-	-	-	-
F2	Softmax	-	4	-	-	-	-	850
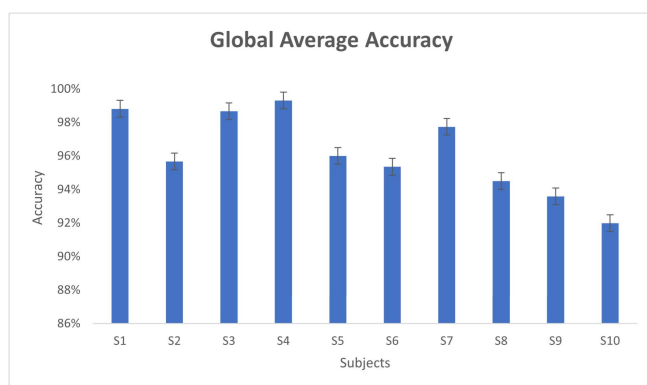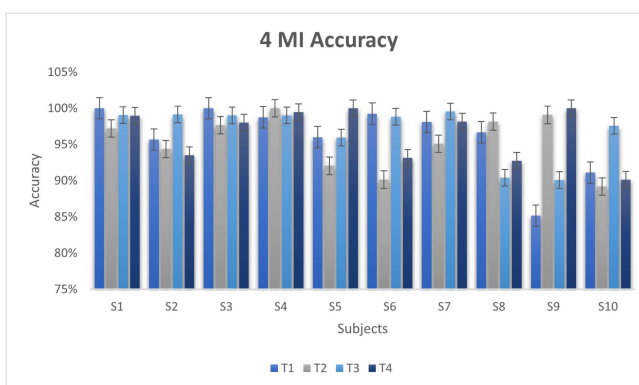
FIGURE 5. The averaged global accuracy of an individual subject across a dataset of ten subjects. (a) A column chart illustrating an individual subject's global averaged accuracy. (b) Column chart of an individual subject's four types of MI accuracy.

to data changes. To ensure that the model was valid, 10% of the dataset was employed as a test set. The training and test sets were then normalised before being sent to CNN for processing.

A. EVALUATION METRICS

The following metrics are used to compare the proposed method performance to that of other approaches: accuracy, F1 score and receiver operating characteristic (ROC) curve.
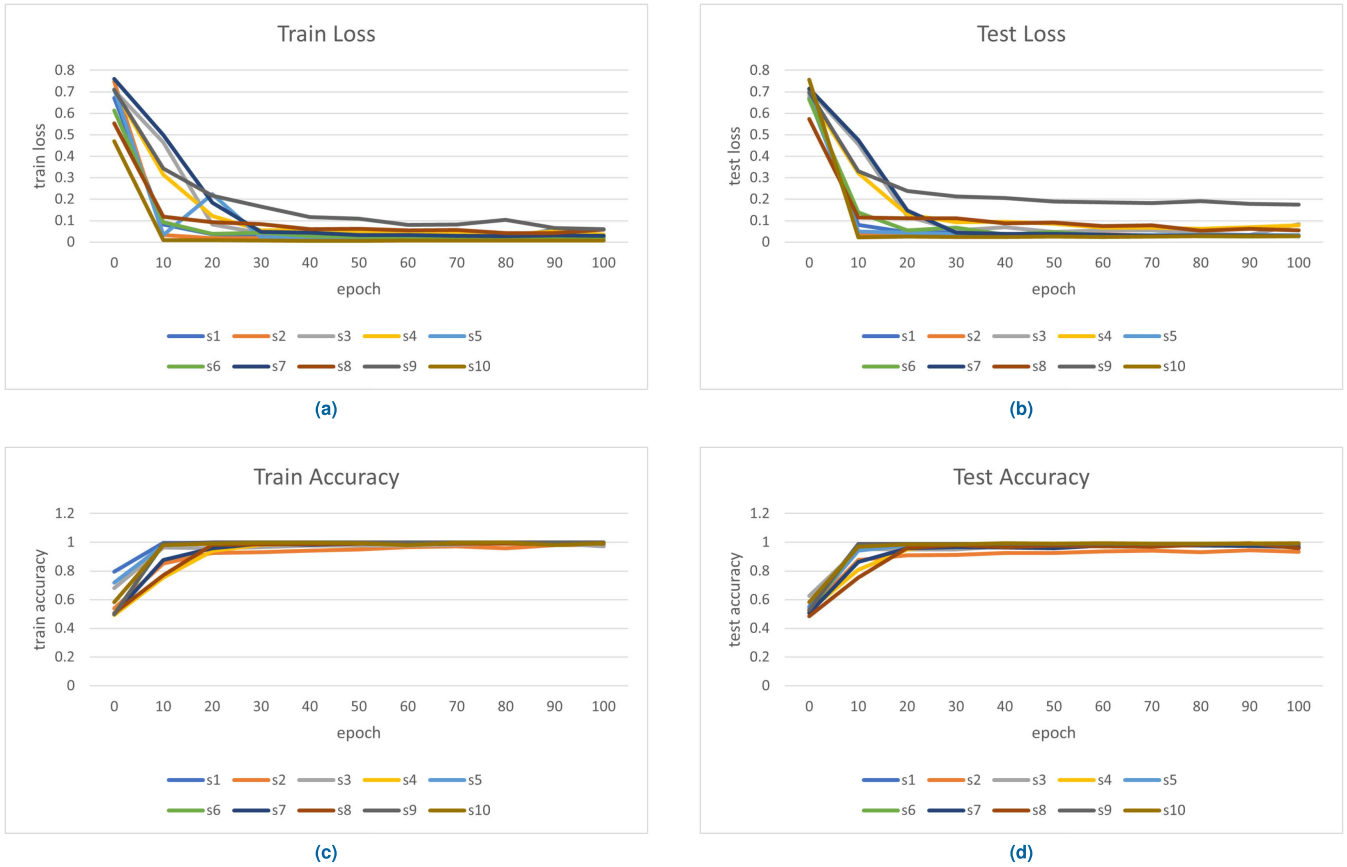
(a)

(b)

(c)

(d)

FIGURE 6. Comparison of classification performance on the physionet dataset. (a) and (b) show training and test loss function curve for 10 subjects. (c) and (d) show training and test accuracy function curve for 10 subjects.

The following formulas are used to calculate these metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (16)$$

where true positive (TP) is the number of positive samples correctly classified and true negative (TN) is the number of negative samples correctly classified. False positive (FP) refers to the number of samples with an original negative label that are classified as positive class, while false negative (FN) refers to the number of positive class samples that are predicted to be negative class.

The AUC criterion is another important criterion for determining the efficiency of a classifier. As shown in Fig 7, the area under the curve (AUC) is the probability that a classifier will rank a set of data at a random chosen positive instance higher than a random chosen negative instance and has a range of 0.5 to 1. The method is more reliable if the AUC is close to 1.0.
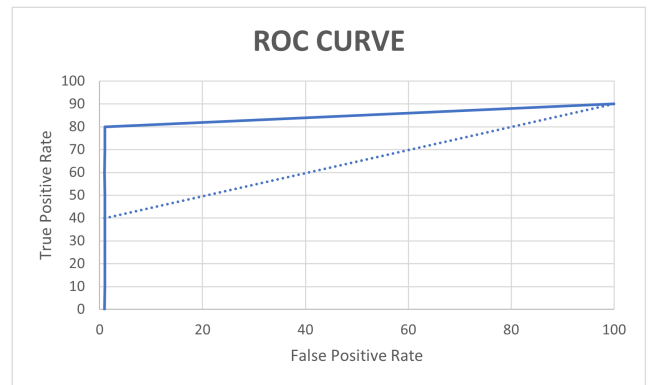
FIGURE 7. ROC curve.

B. CLASSIFICATION ACCURACY

To obtain valid results, all trials of a specific subject were divided into 10 parts, nine for training and one for testing to make sure no data blocks were split between the training and test sets as a result of this.

The model was then trained and tested for accuracy. Data segmentation, training, and testing were all part of the process. Moreover, for each subject, ten cycles were created. Their average was used to get the global averaged accuracy

FIGURE 8. The physionet dataset mean confusion matrices for all subjects.
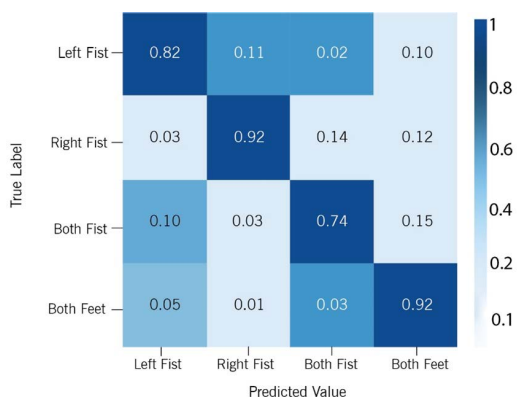
TABLE 2. Performance (ACC and AUC) of the proposed method for 10 subjects classification on physionet dataset.

Subject	ACC	AUC
S1	99.07%	94.24%
S2	97.67%	98.65%
S3	98.67%	94.84%
S4	99.31%	96.28%
S5	95.75%	97.96%
S6	95.52%	96.28%
S7	97.73%	94.86%
S8	97.40%	92.89%
S9	98.58%	99.99%
S10	97.48%	99.08%

TABLE 3. Comparison of results using the Physionet EEG database.

Work	MI tasks	Accuracy	Methods
Ma et al. (2018)	4	68.20%	RNNs
Pinheiro et al. (2018)	4	74.96%	RNA
Dose et al. (2018)	4	80.38%	CNN
Karácsony et al. (2019)	4	76.37%	CNN
Hou et al. (2020)	4	94.54%	ESI+CNN
Lun et al. (2020)	4	95.76%	CNN
This work	4	97.72%	DAE+CNN

of each subject. The overall accuracy (ACC) and area under curve (AUC) for the 10 subjects are shown in Table 2.

The global average accuracy of ten subjects is 96.59%, as shown in Fig 5(a). S4 had the best classification result. It has 98.76% (T1), 100% (T2), 99.01% (T3), and 99.46% (T4) MI accuracies, respectively. S10 has the lowest average accuracy, while MI accuracies are 91.11% (T1), 89.17% (T2), 97.56% (T3), and 90.12% (T4), respectively as shown in Fig 5(b). T2 has the lowest accuracy, indicating that S10 has the worst classification effect on T2. On four MI, the average accuracy of ten subjects is 96.08% (T1), 95.29% (T2), 96.87% (T3), and 96.41% (T4). Both fists and right fist are the best and worst of the four types of MI tasks respectively.

Fig 6 illustrates the loss and accuracy function curves of the physionet dataset for 10 subjects. The convergence of the models can be seen under a variety of conditions. The number of iterations is represented by the abscissa, while the loss and accuracy values are represented by the ordinate. As shown in Fig 6(a) and (b), for the first two epochs, the loss on the training set decreases rapidly. The loss in the test set does not decrease at the same rate as in the training set, but instead remains nearly flat over several epochs. This indicates that our model generalises well to new data. Moreover, In the first two epochs, the accuracy increases rapidly, indicating that the network is learning quickly. After that, the curve flattens, indicating that the model can be trained with fewer epochs as shown in Fig 6(c) and (d).

The mean confusion matrices for all subjects is shown in Fig 8 which show their group-level classification results. The percentage of correct classification is represented by the numbers in the diagonal lines, while the percentage of misclassification is represented by the numbers in the other lines. Both the right hand and both feet had the best MI discrimination.

C. PERFORMANCE COMPARISON

The work of Ma et al. [54], Pinheiro et al. [55], Dose et al. [49], Karácsony et al. [56], Hou et al. [57] and

Lun et al. [58] who performed the same MI tasks on the same dataset were compared to verify the effectiveness of the proposed approach, as shown in Table 3.

Hou et al. [57] applied the Colin27 average brain for the Physionet database, the OpenMEEG toolbox's implementation of the boundary element method (BEM) for a realistic-geometry head model, and extract features by using the Morlet wavelet approach. Its preprocessing procedure is extremely complex. Lun et al. [58], on the other hand, did not perform any preprocessing or feature extraction on the raw data which used as CNN input.

The proposed method outperforms the most effective presented models, demonstrating that DAE improves model decoding performance and CNN improves model generalisation performance. The CNN method is effective and improves the model's generalisation performance in the classification of MI.

While EEGs are frequently low-amplitude and noisy, the patterns are not consistent between the various subjects and patterns. However, the cross-subject model has a high level of accuracy and performs well. In comparison to most previous works, the VAE performs well in terms of removing the noise and then the DAE are generating the features which are fed to the CNN for classification.

The accuracy and stability of high-dimensional BCI classification are improved with this approach while also improving their generalisation ability. As a result, a BCI may become more widely used and not just suitable for one individual.

V. CONCLUSION

This paper presents a novel DAE and CNN classifier for motor imagery classification using EEG signals. EEG signals from ten subjects were recorded while they performed four MI tasks to assess the proposed framework performance. The input images created by EEG signals are first denoised by using the VAE algorithm and then trained for feature extraction by the DAE in this framework. The extracted features are then classified using the CNN model. Using the physionet dataset to compare the results of the DAE-CNN approach to those of other approaches, it is demonstrated that our method produces the most accurate results.

This study's findings demonstrate that the proposed framework is capable of distinguishing between MI tasks within subjects. Moreover, this approach can be further tested with a clinical-grade EEG system, and it should be investigated to see if the number of electrodes can be reduced even further.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain–computer interface," *Nature*, vol. 442, no. 7099, pp. 195–198, Jul. 2006.

[3] K. Stingl, K. U. Bartz-Schmidt, D. Besch, C. K. Chee, C. L. Cottriall, F. Gekeler, M. Groppe, T. L. Jackson, R. E. MacLaren, and A. Koitschev, "Subretinal visual implant alpha IMS–clinical trial interim report," *Vis. Res.*, vol. 111, pp. 149–160, Jun. 2015.

[4] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, Jul. 2006.

[5] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ, USA: Wiley, 2013.

[6] T. Yanagisawa, M. Hirata, Y. Saitoh, T. Goto, H. Kishima, R. Fukuma, H. Yokoi, Y. Kamitani, and T. Yoshimine, "Real-time control of a prosthetic hand using human electrocorticography signals," *J. Neurosurgery*, vol. 114, no. 6, pp. 1715–1722, Jun. 2011.

[7] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, Sep. 2011.

[8] B. Kaur, D. Singh, and P. P. Roy, "A novel framework of EEG-based user identification by analyzing music-listening behavior," *Multimedia Tools Appl.*, vol. 76, no. 24, pp. 25581–25602, Dec. 2017.

[9] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimedia Tools Appl.*, vol. 76, no. 18, pp. 19087–19111, Mar. 2017.

[10] H. Gauba, P. Kumar, P. P. Roy, P. Singh, D. P. Dogra, and B. Raman, "Prediction of advertisement preference by fusing EEG response and sentiment analysis," *Neural Netw.*, vol. 92, pp. 77–88, Aug. 2017.

[11] S.-M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface," *Inf. Sci.*, vol. 178, pp. 1629–1640, Mar. 2008.

[12] G. Reshmi and A. Amal, "Design of a BCI system for piloting a wheelchair using five class MI based EEG," in *Proc. 3rd Int. Conf. Adv. Comput. Commun.*, Aug. 2013, pp. 25–28.

[13] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, and J. Li, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Jul. 2018.

[14] S. Lu, Z. Lu, P. Phillips, S. Wang, and Y. Zhang, "Pathological brain detection in magnetic resonance imaging using combined features and improved extreme learning machines," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1486–1490, 2018.

[15] Z. Zhao, C. Liu, Y. Li, Y. Li, J. Wang, B.-S. Lin, and J. Li, "Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks," *IEEE Access*, vol. 7, pp. 34060–34067, 2019.

[16] X. Li, H. Fan, H. Wang, and L. Wang, "Common spatial patterns combined with phase synchronization information for classification of EEG signals," *Biomed. Signal Process. Control*, vol. 52, pp. 248–256, Jul. 2019.

[17] M. D. Basar, A. D. Duru, and A. Akan, "Emotional state detection based on common spatial patterns of EEG," *Signal, Image Video Process.*, vol. 14, no. 3, pp. 473–481, Apr. 2020.

[18] F. Shiman, E. López-Larraz, A. Sarasola-Sanz, N. Irastorza-Landa, M. Spüler, N. Birbaumer, and A. Ramos-Murguialday, "Classification of different reaching movements from the same limb using EEG," *J. Neural Eng.*, vol. 14, no. 4, 2017, Art. no. 046018.

[19] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018.

[20] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 204–207.

[21] R. Fu, Y. Tian, P. Shi, and T. Bao, "Automatic detection of epileptic seizures in EEG using sparse CSP and Fisher linear discrimination analysis algorithm," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–13, Feb. 2020.

[22] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.

[23] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2017, Art. no. 016003.

[24] H. K. Lee and Y.-S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequecy image," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2018, pp. 906–909.

[25] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, Jan. 2017.

[26] L. Cheng, D. Li, G. Yu, Z. Zhang, X. Li, and S. Yu, "A motor imagery EEG feature extraction method based on energy principal component analysis and deep belief networks," *IEEE Access*, vol. 8, pp. 21453–21472, 2020.

[27] M. A. Rahman, M. Hossain, F. Khanam, M. K. Alam, and M. Ahmad, "Four-class motor imagery EEG signal classification using PCA, wavelet, and two-stage neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, 2019.

[28] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, and A. K. Nandi, "Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction," *Concurrency Comput., Pract. Exper.*, vol. 30, no. 23, p. e4446, Dec. 2018.

[29] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Netw.*, vol. 19, no. 10, pp. 1768–1782, Oct. 2008.

[30] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.

[31] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information," *BMC Bioinf.*, vol. 18, no. S16, pp. 125–137, Dec. 2017.

[32] Y. Park and W. Chung, "Optimal channel selection using correlation coefficient for CSP based EEG classification," *IEEE Access*, vol. 8, pp. 111514–111521, 2020.

[33] S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based OPTICAL predictor," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.

[34] M. R. Islam, T. Tanaka, and M. K. I. Molla, "Multiband tangent space mapping and feature selection for classification of EEG during motor imagery," *J. Neural Eng.*, vol. 15, no. 4, Aug. 2018, Art. no. 046021.

[35] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.

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

IEEE *Access*

[37] H. K. Lee and Y.-S. Choi, "Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface," *Entropy*, vol. 21, no. 12, p. 1199, Dec. 2019.

[38] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarés-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[39] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, and M. Dong, "A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification," *Frontiers Neurosci.*, vol. 13, p. 1275, Nov. 2019.

[40] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. Shamim Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, Dec. 2019.

[41] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.

[42] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[43] R. Alazrai, M. Abuhijleh, H. Alwanni, and M. I. Daoud, "A deep learning framework for decoding motor imagery tasks of the same hand using eeg signals," *IEEE Access*, vol. 7, pp. 109612–109627, 2019.

[44] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019.

[45] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, p. 210, 2019.

[46] O. Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.

[47] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[48] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[49] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018.

[50] J. F. Hwaidi and T. M. Chen, "A noise removal approach from EEG recordings based on variational autoencoders," in *Proc. 13th Int. Conf. Comput. Autom. Eng. (ICCAE)*, Mar. 2021, pp. 19–23.

[51] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.

[52] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. Int. Conf. Artif. Neural Netw.*, Berlin, Heidelberg: Springer, 2010, pp. 92–101.

[53] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8609–8613.

[54] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1903–1906.

[55] O. R. Pinheiro, L. R. G. Alves, and J. R. D. Souza, "EEG signals classification: Motor imagery for driving an intelligent wheelchair," *IEEE Latin Amer. Trans.*, vol. 16, no. 1, pp. 254–259, Jan. 2018.

[56] T. Karácsony, J. P. Hansen, H. K. Iversen, and S. Puthusserypady, "Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback," in *Proc. 10th Augmented Hum. Int. Conf.*, Mar. 2019, pp. 1–8.

[57] Y. Hou, L. Zhou, S. Jia, and X. Lun, "A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016048.

[58] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, "A simplified CNN classification method for MI-EEG via the electrode pairs signals," *Frontiers Hum. Neurosci.*, vol. 14, p. 338, Sep. 2020.

JAMAL F. HWAIDI received the master's degree in computer systems and network engineering from the University of Greenwich, U.K., in 2017. He is currently pursuing the Ph.D. degree in electrical engineering with the Department of Electrical and Electronic Engineering, City, University of London, U.K. He is also working as a Teaching Assistant with the School of Mathematics, Computer Science and Engineering, City, University of London. His research interests include brain-machine interface, deep learning, and EEG signal processing.

THOMAS M. CHEN is currently a Professor in cyber security with the City, University of London. He was previously a Senior Technical Member of Staff with the GTE Laboratories (now Verizon), Waltham, Massachusetts. He joined Southern Methodist University, Dallas, as an Associate Professor, and then Swansea University, Wales, as a Professor in networks. He has served as the Former Editor-in-Chief for *IEEE Communications Magazine*, *IEEE Network,* and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He was a co-recipient of the Fred W. Ellersick Best Paper Award, in 1996. He has written or edited seven books, 41 book chapters, and owns two U.S. patents. His research interests include cyber security, online extremism, and computer networks.

• • •