Biomedical Signal Processing and Control 69 (2021) 102826

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Electroencephalography-based motor imagery classification using temporal convolutional network fusion

Yazeed K. Musallam [a], Nasser I. AlFassam [a], Ghulam Muhammad [a,b,*], Syed Umar Amin [a,b], Mansour Alsulaiman [a,b], Wadood Abdul [a,b], Hamdi Altaheri [a,b], Mohamed A. Bencherif [a,b], Mohammed Algabri [b,c]

[a] Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[b] Center of Smart Robotics Research, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[c] Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:
Motor imagery
Electroencephalography (EEG)
Temporal convolutional network

ABSTRACT

Motor imagery electroencephalography (MI-EEG) signals are generated when a person imagines a task without actually performing it. In recent studies, MI-EEG has been used in the rehabilitation process of paralyzed patients, therefore, decoding MI-EEG signals accurately is an important task, and it is difficult task due to the low signal-to-noise ratio and the variation of brain waves between subjects. Deep learning techniques such as the convolution neural network (CNN) have shown an impact in extracting meaningful features to improve the accuracy of classification. In this paper, we propose TCNet-Fusion, a fixed hyperparameter-based CNN model that utilizes multiple techniques, such as temporal convolutional networks (TCNs), separable convolution, depth-wise convolution, and the fusion of layers. This model outperforms other fixed hyperparameter-based CNN models while remaining similar to those that utilize variable hyperparameter networks, which are networks that change their hyperparameters based on each subject, resulting in higher accuracy than fixed networks. It also uses less memory than variable networks. The EEG signal undergoes two successive 1D convolutions, first along with the time domain, then channel-wise. Then, we obtain an image-like representation, which is fed to the main TCN. During experimentation, the model achieved a classification accuracy of 83.73 % on the four-class MI of the BCI Competition IV-2a dataset, and an accuracy of 94.41 % on the High Gamma Dataset.

1. Introduction

Brain-computer interfaces (BCI) provide a form of communication between the human brain and electronic devices. Research is typically done by recording the neural activities using non-invasive scalp electroencephalography (EEG). Motor imagery (MI) signals are EEG signals that are generated when the subject imagines a movement without actually performing it; MI-BCI helps in rehabilitating humans with impairments or allowing for independence [1]. The sensorimotor cortex area of the brain is the location where the MI-EEG signal is generated. Decoding MI remains a difficult task, as MI is highly dynamic from subject to subject, and it suffers from the low signal-to-noise ratio (SNR) as well as a limited spatial resolution [2]. As a result, designing a model that is robust against low SNR, limited spatial resolution, and other degrading features of EEG acquisition has been a topic of interest in research in recent decades [3].

The BCI research field mainly aims to help paralyzed patients regain control of their limbs and to integrate them into the social environment; a further goal is to allow patients to control prosthetic arms or spell words on a computer screen by only using their thoughts [4–6]. The MI-BCI suffers from the variability of brain signals between subjects, this remains a challenge in MI-BCI when using a global model for all subjects [7–9].

The conventional machine learning algorithms used in MI-BCI decoding and feature extraction are based on handcrafted features. The common spatial pattern (CSP) and filter bank common spatial pattern (FBCSP) are common algorithms used to extract MI features, while classifiers such as linear discriminant analysis (LDA) and support vector machine (SVM) are frequently used in MI-BCI research [1,10,11]. Dimensionality reduction techniques such as principal component

* Corresponding author at: Department of Computer Engineering, College of Computer and information Sciences, King Saud University, Riyadh, Saudi Arabia.
E-mail address: ghulam@ksu.edu.sa (G. Muhammad).

https://doi.org/10.1016/j.bspc.2021.102826
Received 25 March 2021; Received in revised form 7 May 2021; Accepted 24 May 2021
Available online 3 June 2021

analysis (PCA) and independent component analysis (ICA) are used for noise removal [1,11,12].

In recent studies, deep learning (DL) methods have been used in a variety of fields, such as speech recognition, image classification, and natural language processing, and have shown a high impact in extracting complex meaningful features to produce state-of-the-art results for a variety of tasks [13–19]. Recently, conventional neural networks (CNN) outperformed state-of-the-art results when used in image processing applications [20,21]. DL methods have been applied to analyze and classify time-series-related problems; however, CNNs have also shown success with physiological signals, such with as an electrocardiogram (ECG) to detect arrhythmias [22], EEG signals for seizure detection [23], sleep pattern detection [24], and depression [25]. CNNs can extract both local features and general features from the raw EEG, yet it remains a difficult task for CNNs to extract meaningful information due to the low SNR and non-stationarity of EEG signals. Recent studies have shown the success of CNNs in extracting temporal and spatial features from MI-EEG [26].

In the model used in [2], temporal convolution networks (TCNs) were used with the EEGnet architecture to boost accuracy. The main benefit of using TCNs is their capability to extend their receptive field while increasing the number of parameters linearly, as opposed to traditional CNNs, which only increases its field linearly. Another major benefit of using TCNs is that they do not suffer from exploding or vanishing gradient issues in particular, which other time series classification networks do, such as the recurrent neural network (RNN) [2].

This paper focuses on improvements made to the EEG-TCNet model that increased the accuracy of the model while maintaining the compactness of its predecessor. The input EEG signals undergo a 1D convolutional operation in the temporal domain, followed by another 1D channel-wise convolutional operation. After these two sets of 1D convolution operations, we obtain an image-like representation (2D matrix). This image-like representation is then fed into the TCNet. We name the modified model the TCNet-Fusion model. The main contributions of this paper are as follows.

- We propose an efficient TCNet-Fusion model for MI-EEG classification. Modifications to the actual TCN model are done by adding a fusion layer to build rich feature maps to improve the performance of the model.
- An evaluation of the proposed model is conducted using the BCI Competition IV-2a dataset (BCI-IV 2a) [27], with an accuracy of 83.73 % for the within-subjects technique.
- Further benchmarking is conducted using the High Gamma Dataset (HGD) [28], the proposed model achieved an accuracy of 94.41 %.

The paper is organized as follows. In Section 2, we provide a brief literature review of related works on MI-EEG classification techniques and MI datasets. In Section 3, we present our proposed TCNet-Fusion model, and in Section 4 we present the experimental results with a discussion. Lastly, Section 5 concludes the paper.

2. Related literature review

In this section, we first discuss related work. Then, we present two publicly available MI-EEG datasets.

2.1. Related work

Recently, various feature extraction and classification techniques have been used in the field of MI-EEG research. The CSP and other variants, such as the FBCSP, have shown considerable progress in MI classification; FBCSP has won many EEG decoding competitions, including the BCI Competition IV-2a dataset competition [7,29]. On the other hand, DL techniques are used without relying on handcrafted features. Schirrmeister et al. [28] proposed a Deep ConvNet architecture that is able to extract different types of features and can be used as a multi-purpose architecture. The architecture consists of four convolution blocks; the first convolution is performed across time, and the second convolution is performed across space, which acts as an analogue to the CSP spatial filter. The first and second convolution act as one layer without the use of an activation function in between. Moreover, in the same paper [28], ShallowConvNet was introduced; it was designed to be similar to the FBCSP transformation. The first layer performed spatial convolution, and the second was a temporal convolution layer that performed similarly to FBCSP, followed by nonlinearity and average pooling.

Lawhern et al. [26] proposed EEGnet, which is an improved Deep ConvNet and ShallowConvNet architecture. The EEGnet is a general-purpose architecture used in many BCI paradigms that achieves state-of-the-art results and can be trained on small amounts of data. EEGnet used the BCI Competition IV-2a MI dataset by resampling the samples to 128 Hz. The architecture consists of three convolution blocks. The first convolution layer performed convolution in the time domain with a temporal filter at half of the sampling rate, which yielded feature maps of the band-pass frequencies. It then performed a depth-wise convolution that was applied with the filter size of the number of channels (1). This enabled the extraction of spatial filters for each temporal filter and was controlled by the depth parameter D. Then separable convolution was used to reduce the number of parameters and alleviate the relationship between features maps.

Riyad et al. [30] proposed Incep-EEGNet and achieved an accuracy of 74.074 % on BCI-IV 2a. The architecture was based on an inception module and EEGnet. The first and second convolutional layers performed the same way as EEGnet and an inception module with several branches. The inception block increased the number of feature maps, resulting an increase in accuracy. The dataset was also resampled to 128 Hz using a bandpass filter between 1 and 32 Hz, and data cropping augmentation techniques were used to improve the performance.

The feature fusion networks have shown cross-subject improvement in classifying EEG-MI by alleviating overfitting and extracting complex features in a multi-branched CNN by building up composite feature maps concatenated by a fusion layer. One study [31] proposed EEGnet Fusion for a multi-branched convolution neural network, which achieved an accuracy of 84.1 % in cross-subject classification manner using the EEG Motor Movement/Imagery Dataset (eegmmidb) [32]. Each branch in the EEGnet fusion network matched the EEGnet model but differed in the number of filters and kernel size.

Study [33] proposed a channel-projection mixed-scale CNN, which achieved accuracies of 74.6 % on BCI-IV 2a and 93.7 % on HGD. The first block of the network extracted both spatial and temporal information, which was represented as a 1D array, and the second block performed a dilated and regular convolution to produce a rich feature map. The dilated convolution extracted extensive amounts of temporal information and reduced the number of parameters. All feature maps were then concatenated and fed into the classification block. Prior to training, amplitude-perturbation data augmentation was used to increase the number of trials to combat the overfitting problem.

Zhao et al. [34] proposed a multi-branched 3D CNN, which used a 3D representation of the EEG to extract temporal and spatial features simultaneously, resulting in a decrease in the standard deviation among subjects. The architecture consisted of three 3D CNN branches, each having a different size of receptive field, which improved performance. Each 3D CNN branch extracted spatial and temporal features simultaneously. The later layers extracted generic features from local features, and then the branched networks were added. Cropped 3D sliding windows were used to increase the overall performance.

In study [2], a novel temporal convolution network model, EEG-TCNet, was proposed. The network achieved high accuracy with low computational complexity and trainable parameters. The main purpose of this model was to take advantage of its low memory footprint, making it suitable for embedded classifications on

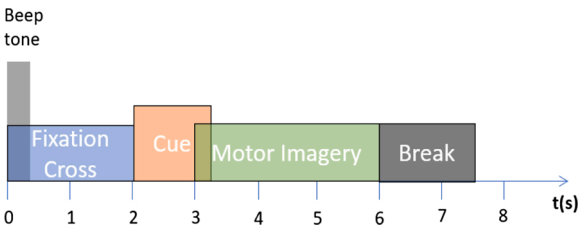

Fig. 1. Timing Paradigm of BCI-2a.

resource-limited devices. The model achieved an accuracy of 77.35 % using fixed parameters. This was improved to 83.84 % by optimizing the network hyper-parameters per subject.

2.2. BCI Competition IV-2a dataset

The BCI Competition IV-2a dataset [27] contains the recordings of nine subjects. The data were collected using 22 EEG electrodes with a sampling frequency of 250 Hz and were then filtered using a bandpass filter between 0.5 Hz and 100 Hz. The dataset contains four different classes of imagined movements, where each class represents a body part: left hand, right hand, feet, and tongue. In addition, three electrooculography (EOG) channels were used to provide information regarding eye movement. Each subject recorded a total of two sessions on different days. The first was used for training while the other was used for testing. Each session contained 288 trials, and trials contaminated with artifacts were excluded. All subjects were seated comfortably in an armchair in front of a computer screen. For each trial, at t = 0, a fixed cross appears on a black screen, alongside a short warning tone to signal the start of a trial. After two seconds (t = 2), a cue appears in the form of an arrow that points in a direction that corresponds to one of the four classes, which are left for left hand, right for right hand, down for feet, and up for tongue. The arrow remains for 1.25 s and prompts the subject to perform the desired class. The subjects were instructed to carry out the desired MI task until the fixed cross disappeared from the screen. A short break of 2 s during which the screen was black ended each trial. The timing scheme of each trial is shown in Fig. 1.

2.3. High Gamma Dataset

The HGD includes more trials than the BCI-IV 2a with around 880 trials in the training set and 160 in the test set with the four classes: left hand, right hand, feet, and rest. The HGD was obtained from 14 subjects (6 female, 2 left-handed, age 27.2 ± 3.6 (mean ± std)) in a controlled environment. Each subject completed the MI movement at their own speed. In addition to the MI, the subjects had to repetitively clench their toes, perform sequential finger-tapping, or relax depending on the direction of a gray arrow that appeared on the screen. These particular movements were chosen because they require little proximal muscular activity while still being complex enough to keep the subjects involved and active. The recordings were acquired in an EEG lab that was optimized for non-invasive data collection utilizing high-frequency movement-related EEG components. The dataset was collected using 128 channels with a sampling frequency of 500 Hz. In each run, 80 arrows appeared for the four classes. With a total of 13 runs performed, each class contained 260 trials with a time window of 4 s for each trial, the inter-trial intervals were 3–4 s [28].

2.4. Temporal convolutional network (TCN)

Recently, a variation of CNN called TCN, which is used in sequencing-related tasks, has been outperforming other recurrent architectures, such as long short-term memory (LSTM) and gated recurrent units (GRUs) [35,36]. TCN allows the extraction of temporal information from sequential data in a powerful manner [2,37].

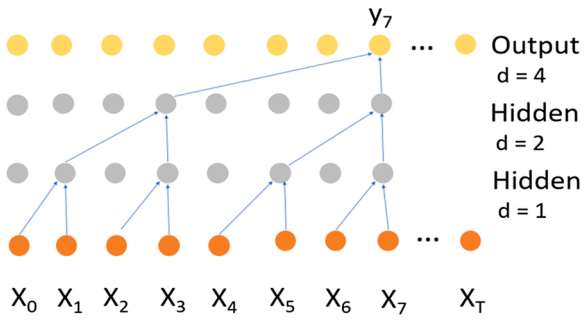

Fig. 2. Dilated causal convolution with kernel size (K_T) = 2 and dilation factors d = [1, 2, 4].

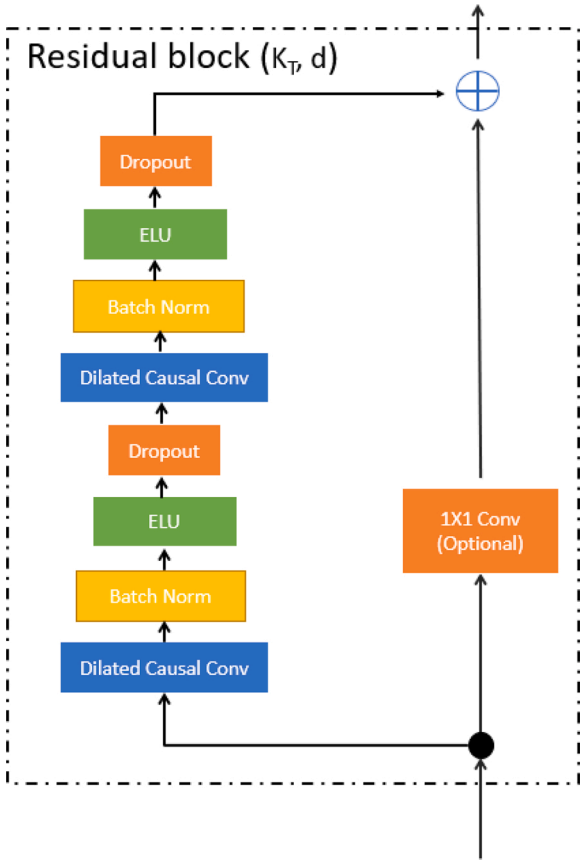

Fig. 3. TCN residual block.

Causal convolution is one type of CNN used to guarantee that no information will leak from the future to the past, meaning the output at time t only depends on the input at time t and earlier. When causal convolutions are used in deep networks, the large kernel size requires a large receptive field size. On the other hand, the dilated convolution allows the receptive field to expand with exponential growth [2,38]. The dilated causal convolutions can learn from a wider range of time-steps, and are controlled by the dilation factor, as illustrated in Fig. 2; however, due to the dilation factor, some time-steps are skipped and only far-away times-step features are extracted [37]. This is a sharp contrast of conventional time-domain signal processing in medical fields [39].

The residual block used in TCN is intended to prevent vanishing and exploding gradient problems in deeper models. Stacking residual blocks allows an increase in the size of the receptive field. The residual block used in EEG-TCNet [2] differs from the one described in [35]. The architectural elements of EEG-TCNet are illustrated in Fig. 3. Each

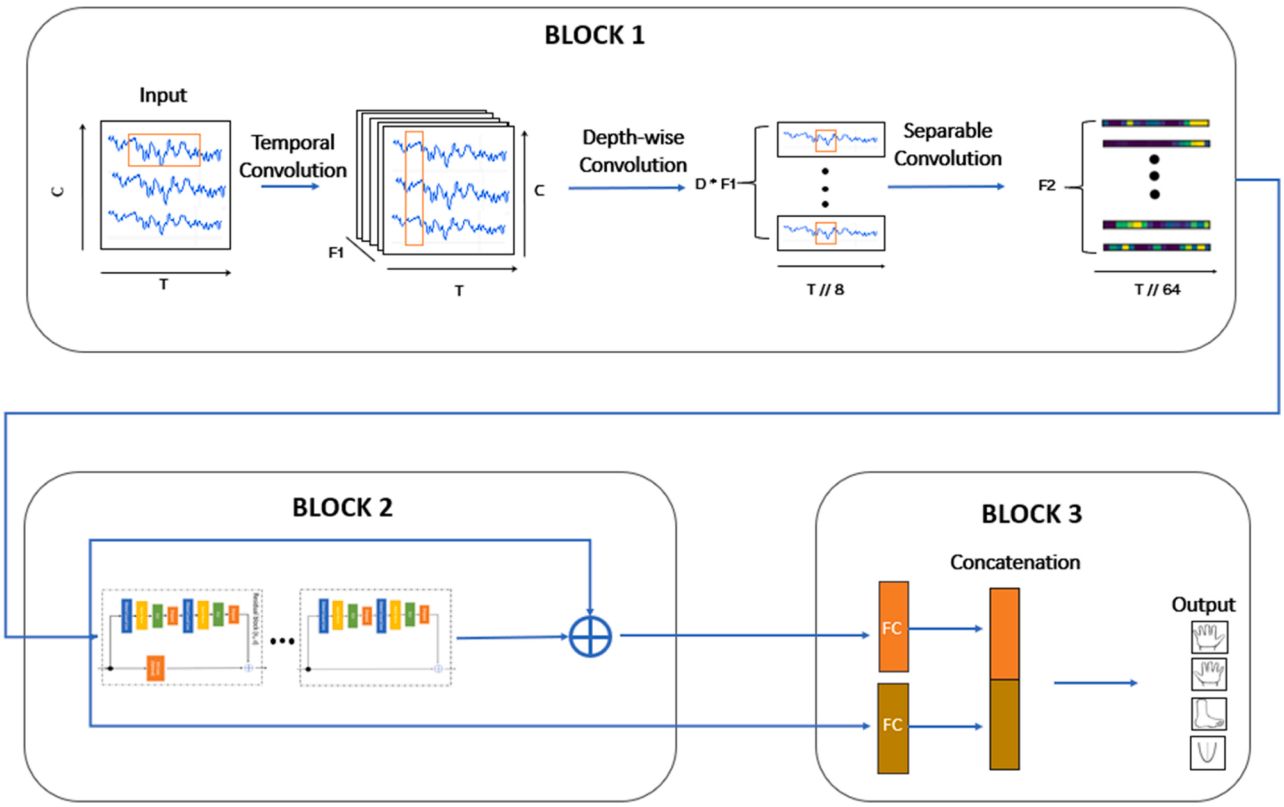

Fig. 4. Proposed TCNet-Fusion model architecture.

residual block consists of dilated causal convolutions, batch normalization (BN), an exponential linear unit (ELU) activation function, regular dropout, and 1×1 convolution added to ensure the addition has the same depth shapes.

3. Methodology

This section describes the data preprocessing, the proposed model TCNet-Fusion description, and the training procedure of the model.

3.1. Data pre-processing

Both the BCI Competition IV dataset 2a (BCI-IV dataset) and HGD were used in this study. The BCI IV 2a was recorded using 22 electrodes with a sampling frequency of 250 Hz. Each trial contained 0.5 s pre-cue onsets to the end of the trial, therefore the total length of one trial is 4.5 s for $(250 \times 4.5=)1125$ samples. Standardization was applied to each channel and no further prepossessing was used. Each trial had the shape of a matrix of dimension (22,1125).

The HGD was originally sampled at 500 Hz, and then the data were resampled to 250 Hz. Moreover, the number of channels was reduced from 128 to 44 to reduce unnecessary information. Furthermore, the pre-processing step used the procedure reported in [31]. A trial length of 4.5 s was used, resulting in 1125 samples for each trial. The trial matrix had the dimensions (44,1125). Standardization was used for each channel, and no additional filters were applied.

3.2. Model description

The proposed model, TCNet-Fusion, was based on EEGnet and EEG-TCNet. The motivation behind the model was to further improve the accuracy of EEG-TCNet, without the need for employing a variable network. This helps the robustness of the model, as the hyper-parameters of the model do not need to be changed for each subject. The architecture of the proposed model is composed of three main parts. The first part is similar to EEGnet and is comprised of three convolution layers: temporal convolution, depth-wise convolution, and separable convolution. The temporal convolution allows different bandpass frequencies to be learned. Depth-wise convolution is used to extract spatial features for each temporal filter. Separable convolution is applied to alleviate the relationship between feature maps and reduce the number of parameters. In the second part, TCN is applied to extract additional temporal features. Concatenation is then applied between the output of the TCN and the EEGnet blocks to alleviate feature loss and build a rich feature map. The last part flattens and concatenates the outputs of EEGnet, and the previous concatenation layer builds up a fused feature layer. Furthermore, the fused layer is fed into a classification block with the SoftMax activation function (Fig. 4).

The inputs (the MI-EEG trials) are fed into the temporal convolution layer with a kernel of size $(1, K_E)$ and F_1 filters. The temporal convolution allows the learning of different bands of frequencies. The temporal convolution is followed by a depth-wise convolution with a kernel size of $(C,1)$ with a depth parameter of D to extract frequency-specific spatial filters and control the number of spatial filters to be learned for each temporal filter, respectively. After the depth-wise convolution, separable convolutions with a size of (1,16) and F_2 point-wise filters are applied. Separable convolutions are simply a combination of depth-wise convolution and point-wise convolution. The separable convolutions allow mixing between feature maps and learning of the temporal summaries of each feature map. Furthermore, an average pooling layer is applied after the depth-wise and separable convolutions with a size of (1,8), to reduce the number of parameters while keeping the important information.

After each convolutional layer, BN is applied to speed up the training and regularization of the model. Then, an ELU activation function is used [40], as it has experimentally shown better performance than the rectified linear unit (ReLU) activation function. A regular dropout function is used after the separable convolution to regularize the

Table 1
Detailed architecture of the proposed model, where C = number of channels, T = number of samples, F_1 = number of temporal filters, F_2 = number of pointwise filters, D = depth multiplier, K_E = kernel size of EEGnet first convolution, K_T = kernel size in TCN block, p_e = dropout rate of EEGnet, p_t = dropout rate in TCN block, F_T = number of convolution filters in TCN block.

Layer	Layer type	Maps	Kernel size	Output	Options
Input	Input	–	–	(1, C, T)	–
C1	Conv2D	F_1	(1, K_E)	(F_1, C, T)	padding = same
BN1	BatchNorm	–	–	(F_1, C, T)	
C2	DepthwiseConv2D	D * F_1	(C, 1)	(D*F_1, 1, T)	padding = valid depth = D max norm = 1
BN2	BatchNorm	–	–	(D*F_1, 1, T)	–
A1	Activation	–	–	(D*F_1, 1, T)	Function = ELU
P1	AveragePooling2D	–	(1, 8)	(D*F_1, 1, T // 8)	
D1	Dropout	–	–	(D*F_1, 1, T // 8)	Dropout Rate = p_e
C3	SeparableConv2D	F_2	(1, 16)	(F_2, 1, T // 8)	Padding = same
B2	BatchNorm	–	–	(F_2, 1, T // 8)	–
A2	Activation	–	–	(F_2, 1, T // 8)	Function = ELU
P2	AveragePooling2D	–	(1, 8)	(F_2, 1, T // 64)	–
D2	Dropout	–	–	(F_2, 1, T // 64)	Dropout rate = p_e
TC	TCN	F_T	K_T	(F_T, T // 64)	Dropout rate = p_t Activation function = ELU
CON1	Concatenation	–	–	((F_T + F_2), T // 64)	Concatenate: (D2 & TC)
FC1	Flatten	–	–	(F_2 * T // 64)	Flatten: D1
FC2	Flatten	–	–	((F_T + F_2) * T // 64)	Flatten: CON1
CON2	Concatenation	–	–	(((F_T + F_2) * T // 64) + (F_2 * T // 64))	Concatenate: (FC1 & FC2)
Dense	Dense	–	–	N	max norm = 0.25
SoftMax	SoftMax	–	–	N	–

network.

Following the separable convolutions, a TCN layer is applied to learn more temporal information with *L* stacked residual blocks. Each residual block, as described in [1], contains dilated causal convolutions, BN, an ELU activation function, regular dropout, and 1×1 convolution, as shown in Fig. 3. The increase in dilation factor in the dilated causal convolutions causes the adjacent time-steps to be skipped, and therefore reduces the temporal information that is extracted [37]. Later, the outputs of the separable convolutions and the TCN block are concatenated to build a rich feature map (Table 1).

The input to the TCN can be shown as images because after the temporal and the depth-wise convolutions and the separable convolutions, the signal can be represented by a 2D matrix. Fig. 5 shows four sample images, one for each class.

The outputs of the previously concatenated layer and the separable convolution layer are then flattened and fused to produce a fused feature vector. Utilizing features from the fused vector enhances the performance of the model.

3.3. Training procedure

Both mental and physical states may differ from subject to subject. In the EEG-MI research field, two methods are adopted for classifying MI: within-subject and cross-subject. The within-subject technique, where the training and the testing of the model are done on sessions recorded for the same subject but using different samples, has provided high accuracy in recent studies [1]. The cross-subject technique is used to train the model on all the subjects, leaving out only one subject for testing, then repeating over the remaining subjects; however, the cross-subject technique remains a challenge due to the dynamic nature of the brain waves of each subject [1,31]. In this study, the within-subject technique is employed for both BCI-IV 2a dataset and HGD using the proposed model. For both datasets, one session is used for training and the other session is used in the testing phase. For both datasets, global parameters are used for all subjects in the proposed model, as shown in Table 2.

Table 2
Global hyper-parameters used for all subjects.

Hyper-Parameter	value
Kernel size in EEGnet block (K_E)	32
Dropout rate in EEGnet (p_e)	0.3
#Temporal filters in EEGnet (F_1)	24
# of residual blocks (L)	2
Kernel size in TCN block (K_T)	4
# Filters in TCN block (F_T)	12
Depth multiplier (D)	2
Dropout rate in TCN block (p_t)	0.3
Standardize data	True

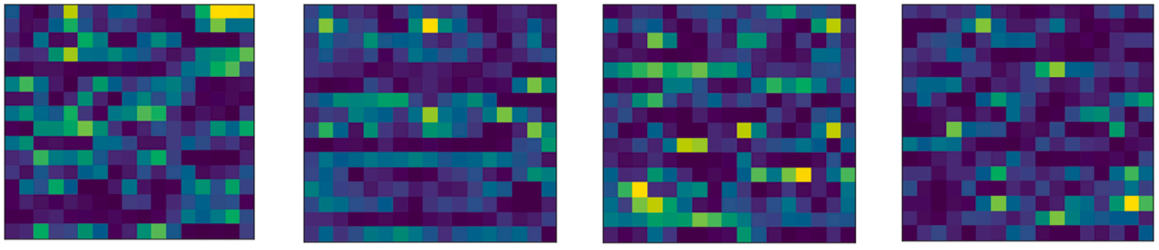

Fig. 5. Image representation of EEG signals after the temporal and depth-wise convolutions and separable convolutions. The images shown are samples of four classes.

Table 3
Classification accuracy (%) on the MI BCI IV-2a dataset.

Subjects	Shallow ConvNet [2]	EEG-TCNet (Fixed) [2]	Incep-EEGNet [30]	Lu et al. [43]	EEGnet (Variable) [2,26]	Liu et al. [41]	Anwar et al. [42]	EEG-TCNet (Variable) [2]	(Proposed Model) TCNet Fusion
1	79.51	85.77	78.47	87.84	86.48	82.99	83.94	89.32	90.74
2	56.25	65.02	52.77	73.95	61.84	56.25	71.23	72.44	70.67
3	88.89	94.51	89.93	83.79	93.41	93.06	87.15	97.44	95.23
4	80.90	64.91	66.66	72.65	73.25	84.03	84.10	75.87	76.75
5	57.29	75.36	61.11	76.66	76.81	68.03	86.40	83.69	82.24
6	53.82	61.40	60.41	67.29	59.07	58.34	78.23	70.69	68.83
7	91.67	87.36	90.62	73.37	90.25	88.20	77.14	93.14	94.22
8	81.25	83.76	82.29	79.60	87.45	88.20	82.56	86.71	88.92
9	79.17	78.03	84.37	77.42	82.95	86.81	79.14	85.23	85.98
Average	74.31	77.37	74.07	76.62	79.06	78.51	81.09	83.84	83.73
Standard deviation	14.54	11.57	14.06	6.18	12.28	13.83	5.11	9.20	9.79

In the training phase, at the end of each epoch, a callback is used to save the best model weights, based on the best accuracy so far, and then the best saved model is loaded during the test phase. The model is trained for 1000 epochs with a batch size of 64, and the learning rate is set to 0.0009. An Adam optimizer was used with a cross-entropy error function.

4. Experiments

The experiments were conducted using the TensorFlow framework. Subject-specific experiments were conducted on both the BCI Competition IV-2a dataset and HGD. This section presents the experimental results with a comparison of state-of-the-art methods. Different performance metrics are formulated. Then, the results using two datasets are described.

4.1. Performance metrics

The model was evaluated using multiple metrics, the first of which is the classification accuracy, which is calculated using Eq. (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP and TN are true positive and true negative, respectively, and FP and FN are false positive and false negative, respectively.

The second metric is Cohen's κ-score, which is calculated using Eq. (2):

$$k = \frac{P_0 - P_e}{1 - P_e} \tag{2}$$

where P_0 is the accuracy of the model, and P_e represents the probability or accuracy of a random guess.

The standard deviation was calculated for both the classification accuracy and kappa values.

Furthermore, a multiply-accumulate operations (MAC) metric for different CNN layers can be calculated, as in Eq. (3) for a 1D convolution layer, Eq. (4) for a 2D convolution layer, and Eq. (5) and Eq. (6) for depth-wise and separable convolutions, respectively.

$$\text{1D Convolution} = K1 * Cin * Cout * Wout \tag{3}$$

$$\text{2D Convolution} = K1 * K2 * Cin * Cout * Hout * Wout \tag{4}$$

$$\text{Depth} - \text{wise Convolution} = K1 * K2 * Cin * D * Hout * Wout \tag{5}$$

$$\text{Separable Convolution} = (K1 * K2 + Cout) * Cin * Hout * Wout \tag{6}$$

Information density is calculated as the ratio of the accuracy (between 0 and 1) and # of parameters in millions, as in Eq. (7).

$$\text{Information density} = \frac{Accuracy}{\#\ of\ paramaters\ (M)} \tag{7}$$

The last metric is the F1 score, which requires the calculation of both the precision and the recall obtained from Eqs. (8) and (9), respectively; these are then used in Eq. (10):

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

4.2. Results using BCI Competition IV-2a dataset

In the BCI Competition IV-2a dataset, the proposed model was trained on dataset "T" and tested on dataset "E," which are two sessions conducted on different days, and the experiments were conducted using the subject-specific technique.

The proposed model was compared to state-of-the-art MI-EEG classification methods using multiple metrics, which were classification

Table 4
κ-score on the MI BCI IV-2a dataset.

Subjects	Shallow ConvNet [2]	EEG-TCNet (Fixed) [2]	Incep-EEGNet [30]	Lu et al. [43]	EEGnet (Variable) [2, 26]	EEG-TCNet (Variable) [2]	(Proposed Model) TCNet Fusion
1	0.73	0.81	0.71	0.84	0.82	0.86	0.87
2	0.42	0.53	0.37	0.65	0.49	0.63	0.60
3	0.85	0.93	0.86	0.78	0.91	0.97	0.93
4	0.75	0.53	0.55	0.64	0.64	0.68	0.68
5	0.43	0.67	0.48	0.65	0.69	0.78	0.76
6	0.38	0.49	0.47	0.56	0.45	0.61	0.58
7	0.89	0.83	0.87	0.64	0.87	0.91	0.92
8	0.75	0.78	0.76	0.73	0.83	0.82	0.85
9	0.72	0.71	0.79	0.70	0.77	0.80	0.81
Average	0.66	0.70	0.65	0.69	0.72	0.78	0.78

Table 5
Precision and recall scores on the MI BCI IV-2A dataset.

Metrics/Subjects		1	2	3	4	5	6	7	8	9	Average
Precision	LH	97.10	68.33	98.43	79.24	81.25	71.11	98.41	92.64	90.76	86.36
	RH	97.14	59.25	97.22	72.72	76.54	68.33	88.15	93.84	86.79	82.22
	F	88.88	75.90	89.18	73.61	89.09	76.74	98.52	83.33	84.28	84.39
	Tou	81.01	81.35	96.82	83.33	85	62.68	92.85	86.11	82.89	83.56
Recall	LH	94.36	57.74	94.02	71.18	92.85	60.37	87.32	95.45	90.76	82.67
	RH	97.14	67.60	100	70.17	95.38	74.54	97.10	89.70	70.76	84.71
	F	81.15	91.30	97.05	89.83	68.05	61.11	94.36	79.71	85.50	83.11
	Tou	90.14	66.66	89.70	75.47	73.91	79.24	98.48	91.17	96.92	84.63
F1-Score	Average	90.75	70.48	95.21	76.6	81.95	68.66	94.23	88.95	85.74	83.61

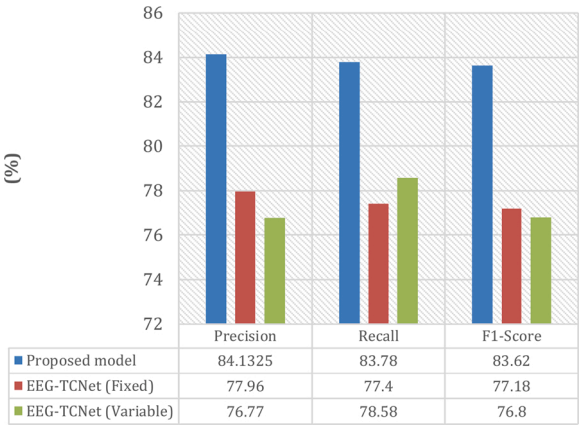

Fig. 6. Average precision, recall, and F1-Score.

accuracy, Cohen's κ-score, precision, recall, F1-score, and number of parameters.

Table 3 summarizes the accuracy of each subject for various of state-of-the-art MI-EEG methods conducted on the BCI Competition IV-2a dataset using a subject-specific manner. The methods that are compared are the proposed TCNet fusion, variable EEGnet [2,26], both fixed and variable EEG-TCNet [2], Shallow ConvNet [2], Liu et al. [41], Anwar et el. [42], Lu et al. [43] and Incep-EEGNet [30]. The proposed model achieved an accuracy of 83.73 %, which was higher than the accuracies obtained by other fixed variable models by at least 2.64 %. Compared with the variable EEG-TCNet model, which achieved an accuracy of 83.84 %, the accuracy of the proposed model was only 0.11 % less, while the model used fixed parameters for all subjects. The proposed model enhanced the accuracies for some subjects in the EEG-TCNet variable network, i.e., Subject 1, Subject 7, and Subject 8; this shows that the proposed model is as good as a variable network, while still using fixed hyperparameters.

The proposed model had better accuracy than fixed networks but a worse standard deviation than the method used by Anwar et al.

Table 4 shows the κ-score comparison for the aforementioned models. The table shows that the proposed model is as good as the variable EEG-TCNet networks, which had a κ-score of 0.78 and was higher (κ-score) than other fixed networks by at least 0.08.

Table 5 summarizes the precision and the recall for each class of the proposed model per subject, as well as the F1-scores of the average precision and recall values. Fig. 6 illustrates the average precision, recall, and F1-Score of the proposed model, fixed EEG-TCNet, and variable EEG-TCNet. The figure shows that the proposed model precision is higher than both fixed and variable networks by at least 6.17 %. Moreover, the recall and F1-scores are higher by at least 6.38 % and 6.82 %, respectively. This indicates the features from different layers using the fusion layer improved the performance of the system.

Fig. 7 presents the confusion matrices of the proposed model for each subject.

Table 6 shows the number of parameters, average MACs, required memory, average accuracy and information density for the proposed model, the fixed and variable EEG-TCNet, Anwar et al., CNN++ [44], and variable EEGnet. In terms of the number of parameters, the proposed model has fewer than variable TCNet by 292,000 parameters, but more than most fixed variable networks. Therefore, the proposed model requires less computational power than the variable TCNet, while still maintaining high accuracy. The proposed model has a higher average MAC value than most variable and fixed models. This is due to the increased number of filters present in the proposed model, meaning the proposed model is unable to perform well with limited-resource devices, such as embedded systems, compared with TCNet. In terms of memory requirements, the proposed model requires less memory than the Anwar et al. method and EEGnet, but more memory than the variable TCNet. This is due to the increase in the F_1 value when compared with TCNet, which uses a smaller F_1 value. For further analysis, the information density metric is used to evaluate the balance between accuracy and the number of parameters. The information density is the ratio of the accuracy (between 0 and 1) and the number of parameters in millions [45, 46]. Table 6 shows that the fixed EEG-TCNet has the highest information density, and the proposed model has a higher information density than the variable EEG-TCNet by 6.73.

4.3. Results using HGD

Table 7 summarizes the precision, recall, F1 score, accuracy, and κ-score for each subject of the HGD. The proposed model achieved an average accuracy of 94.41 %, with a κ-score of 0.92.

5. Conclusion

The TCNet-Fusion model was proposed, which utilized features from different layers of a TCN by using a fusion layer. The intention behind the model was to use global parameters for all subjects, thus outperforming fixed hyperparameter modern models, while remaining similar to models that utilize variable hyperparameter networks for each subject, thus achieving higher accuracy than fixed networks while using less memory than variable networks. The experiments were performed on two public datasets, the BCI Competition IV-2a dataset and the HGD. The within-subject technique was used in the experiment and global hyper-parameters were used for all subjects in both datasets. The average classification accuracy for the BCI-IV 2a was 83.73 % on the four-class MI set, and the average accuracy on the HGD was 94.41 %. For future works, we wish to continue refining the accuracy of BCI-MI classification models and to create models that can be used in advanced BCI systems.

CRediT authorship contribution statement

Yazeed K. Musallam: Methodology, Software, Writing - original draft. **Nasser I. AlFassam:** Methodology, Software, Writing - original draft. **Ghulam Muhammad:** Conceptualization, Supervision, Funding acquisition. **Syed Umar Amin:** Data curation, Validation. **Mansour**

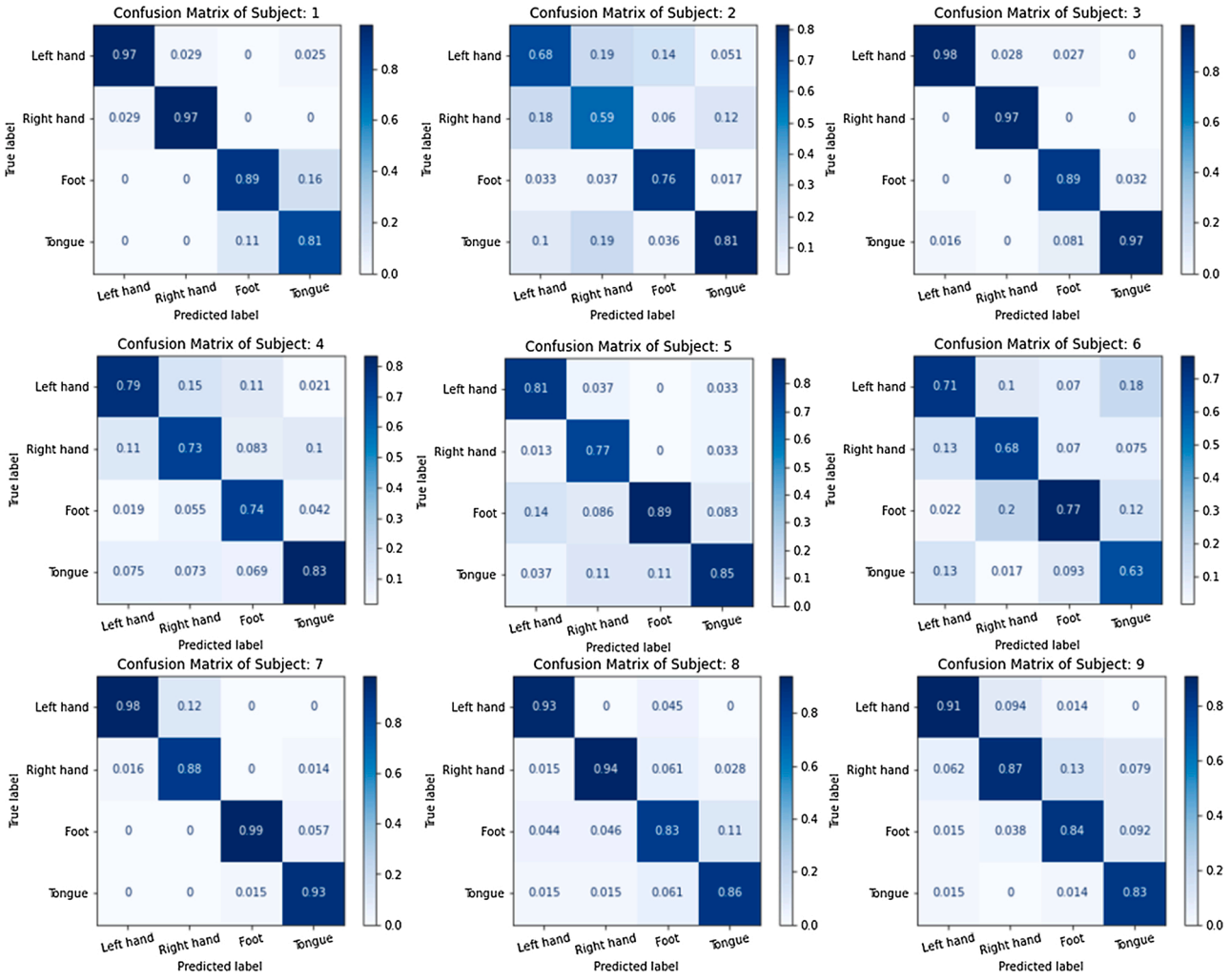

Fig. 7. Confusion matrix for all 9 subjects of the BCI-IV 2a dataset.

Table 6
Number of parameters, MAC, memory requirement, and information density of the proposed model compared with a variety of state-of-the-art methods.

Method	# of Parameters; K = Kilo, M = Million	Average MACs; M = Million	Memory requirement	Average Accuracy	Information density
EEG-TCNet (Fixed) [2]	4.27K	6.8M	396 KB	77.34	181.12
EEGnet (Variable) [2,26]	15.6K	42.6M	1584 KB	79.02	50.65
EEG-TCNet (Variable) [2]	20.5K	12.1M	792 KB	83.84	40.89
CNN++ [44]	220K	18.2M	499 KB	81.10	3.68
Anwar et al. [42] *	4.4M	19097.37M	9.36 MB	81.09	0.18
Proposed Model	17.58K	20.69M	1188 KB	83.73	47.62

* Reproduced.

Table 7
Precision, recall, accuracy, and κ-Score on the HGD.

Metrics/Subjects		1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
Precision	LH	97.56	97.50	100	95.12	100	97.50	90.24	92.30	95.23	97.36	68.42	100	92.68	94.73	94.18
	RH	94.87	100	100	100	95.23	100	89.74	90.24	97.50	90.90	73.80	100	97.36	84.09	93.83
	F	92.68	92.85	100	100	97.56	97.43	100	100	100	85.36	90.69	86.95	100	66.66	93.58
	Re	94.87	95.00	97.56	93.02	100	97.56	100	100	100	94.59	94.59	97.14	97.56	100	97.27
Recall	LH	100	97.50	100	97.50	95.00	97.50	92.50	90.00	100	92.50	65.00	97.50	97.43	45.00	90.53
	RH	92.50	95.00	100	92.50	100	100	89.74	92.50	97.50	100	77.50	100	92.50	92.50	94.44
	F	95.00	97.50	97.50	97.50	100	95.00	100	100	95.00	87.50	97.50	100	97.50	95.00	96.78
	Re	92.50	95.00	100	100	97.50	100	97.50	100	100	87.50	87.50	85.00	100	100	95.89
F1 Score		94.98	96.26	99.37	96.86	98.12	98.11	94.95	95.62	98.12	91.85	81.79	95.60	96.84	81.86	94.30
Accuracy		95.00	96.25	99.37	96.87	98.12	98.12	94.96	95.62	98.12	91.87	81.87	95.62	96.85	83.12	94.41
κ-Score		0.93	0.95	0.99	0.95	0.97	0.97	0.93	0.94	0.97	0.89	0.75	0.94	0.95	0.77	0.92

Alsulaiman: Investigation. **Wadood Abdul:** Writing - review & editing. **Hamdi Altaheri:** Data curation. **Mohamed A. Bencherif:** Investigation. **Mohammed Algabri:** Visualization.

Acknowledgment

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (DRI-KSU-1354).

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Mekhtiche, M.S. Hossain, Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, Future Gener. Comput. Syst. 101 (2019) 542–554.
[2] T.M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, L. Benini, EEG-TCNet: an accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 2958–2965.
[3] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Bencherif, M.S. Hossain, Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification, IEEE Access 7 (December (1)) (2019) 18940–18950.
[4] U. Chaudhary, N. Birbaumer, A. Ramos-Murguialday, Brain–computer interfaces for communication and rehabilitation, Nat. Rev. Neurol. 12 (9) (2016) 513.
[5] M. Kaya, M.K. Binli, E. Ozbay, H. Yanar, Y. Mishchenko, A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces, Sci. Data 5 (1) (2018) 1–16.
[6] P. Wang, A. Jiang, X. Liu, et al., LSTM-based EEG classification in motor imagery tasks, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (11) (2018) 2086–2095.
[7] K.K. Ang, Z.Y. Chin, C. Wang, et al., Filter bank common spatial pattern algorithm on BCI competition iv datasets 2a and 2b, Front. Neurosci. 6 (2012) 39.
[8] B. Blankertz, G. Dornhege, M. Krauledat, et al., The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects, NeuroImage 37 (2) (2007) 539–550.
[9] F. Lotte, C. Guan, K.K. Ang, Comparison of designs towards a subject-independent brain-computer interface based on motor imagery, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 2009, pp. 4543–4546.
[10] Y. Ma, X. Ding, Q. She, et al., Classification of motor imagery EEG signals with support vector machines and particle swarm optimization, Comput. Math. Methods Med. 2016 (2016). Article ID 4941235.
[11] W. Qiao, X. Bi, Deep spatial-temporal neural network for classification of EEG-based motor imagery, Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science (2019) 265–272.
[12] K. Bhimraj, R.J. Haddad, Autonomous noise removal from EEG signals using independent component analysis, in: SoutheastCon 2017, Concord, NC, USA, 2017, pp. 1–6.
[13] A. Al-nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, Investigation of voice pathology detection and classification on different frequency regions using correlation functions, J. Voice 31 (1) (2017) 3–15.
[14] G. Muhammad, M.S. Hossain, N. Kumar, EEG-based pathology detection for home health monitoring, IEEE J. Sel. Areas Commun. 39 (February (2)) (2021) 603–610.
[15] I.D. Falco, et al., Deep neural network hyper-parameter setting for classification of obstructive sleep apnea episodes, in: 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, 2018, pp. 01187–01192.
[16] I.D. Falco, et al., Evolution-based configuration optimization of a Deep Neural Network for the classification of Obstructive Sleep Apnea episodes, Future Gener. Comput. Syst. 98 (September) (2019) 377–391.
[17] F. Alshehri, G. Muhammad, A comprehensive survey of the internet of things (IoT) and AI-based smart healthcare, IEEE Access 9 (January) (2021) 3660–3678.
[18] G. Muhammad, M.F. Alhamid, X. Long, Computing and processing on the edge: smart pathology detection for connected healthcare, IEEE Netw. 33 (November-December (6)) (2019) 44–49.
[19] G. Sannino, I.D. Falco, G.D. Pietro, Non-invasive risk stratification of hypertension: a systematic comparison of machine learning algorithms, J. Sens. Actuator Netw. 9 (3) (2020) 34.
[20] G. Muhammad, M.S. Hossain, COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images, Inf. Fusion 72 (March) (2021) 80–88.
[21] G. Muhammad, S.M.M. Rahman, A. Alelaiwi, A. Alamri, Smart health solution integrating IoT and cloud: a case study of voice pathology monitoring, IEEE Commun. Mag. 55 (January (1)) (2017) 69–73.
[22] U.R. Acharya, H. Fujita, O.S. Lih, et al., Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inf. Sci. 405 (2017) 81–90.
[23] A. Emami, N. Kunii, T. Matsuo, et al., Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images, Neuroimage Clin. 22 (2019), 101684.
[24] F. Li, R. Yan, R. Mahini, et al., End-to-end sleep staging using convolutional neural network in raw single-channel EEG, Biomed. Signal Process. Control 63 (2021), 102203.
[25] G. Sharma, A. Parashar, A.M. Joshi, DepHNN: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression, Biomed. Signal Process. Control 66 (2021), 102393.
[26] V.J. Lawhern, A.J. Solon, N.R. Waytowich, et al., EEGnet: a compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural Eng. 15 (5) (2018), 056013.
[27] C. Brunner, R. Leeb, G. Muller-Putz, A. Schlogl, G. Pfurtscheller, BCI Competition 2008–Graz Data Set A and B, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, 2008, pp. 136–142.
[28] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, et al., Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (2017) 5391–5420.
[29] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 2008, pp. 2390–2397.
[30] M. Riyad, M. Khalil, A. Adib, Incep-eegnet: a convnet for motor imagery decoding, in: International Conference on Image and Signal Processing, Springer, 2020, pp. 103–111.
[31] K. Roots, Y. Muhammad, N. Muhammad, Fusion convolutional neural network for cross-subject EEG motor imagery classification, Computers 9 (3) (2020) 72.
[32] G. Schalk, D.J. McFarland, T. Hinterberger, et al., BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043.
[33] Y. Li, X.-R. Zhang, B. Zhang, et al., A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (6) (2019) 1170–1180.
[34] X. Zhao, H. Zhang, G. Zhu, et al., A multi-branch 3d convolutional neural network for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (10) (2019) 2164–2177.
[35] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271 (2018).
[36] N. Lu, T. Yin, X. Jing, Deep learning solutions for motor imagery classification: a comparison study, in: 2020 8th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea (South), 2020, pp. 1–6.
[37] T.N. Duc, C.T. Minh, T.P. Xuan, E. Kamioka, Convolutional neural networks for continuous qoe prediction in video streaming services, IEEE Access 8 (2020) 116268–116278.
[38] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).
[39] G. Muhammad, T.A. Mesallam, K.H. Malki, et al., Formant analysis in dysphonic patients and automatic Arabic digit speech recognition, Biomed. Eng. Online 10 (2011) 41.
[40] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289 (2015).
[41] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, F. Lin, Parallel spatial–temporal self-attention CNN-Based motor imagery classification for BCI, Front. Neurosci. 14 (2020), 587520.
[42] A.M. Anwar, A.M. Eldeib, EEG signal classification using convolutional neural networks on combined spatial and temporal dimensions for BCI systems, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 434–437.
[43] P. Lu, N. Gao, Z. Lu, J. Yang, O. Bai, Q. Li, Combined CNN and LSTM for motor imagery classification, in: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1–6.
[44] Y. Zhao, S. Yao, S. Hu, S. Chang, R. Ganti, M. Srivatsa, S. Li, T. Abdelzaher, On the improvement of classifying EEG recordings using neural networks, in: 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 1709–1711.
[45] G. Muhammad, M.S. Hossain, A. Yassine, Tree-based deep networks for edge devices, IEEE Trans. Industr. Inform. 16 (3) (2020) 2022–2028.
[46] M.S. Hossain, G. Muhammad, Emotion-aware connected healthcare big data towards 5G, IEEE Internet Things J. 5 (August (4)) (2018) 2399–2406.