IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 June 2022, accepted 3 July 2022, date of publication 13 July 2022, date of current version 19 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190502

RESEARCH ARTICLE

Automated Detection of Major Depressive Disorder With EEG Signals: A Time Series Classification Using Deep Learning

ALIREZA RAFIEI[1], RASOUL ZAHEDIFAR[2], CHIRANJIBI SITAULA[3], AND FAEZEH MARZBANRAD[3], (Senior Member, IEEE)

[1]Department of Mechatronics Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran 1417935840, Iran
[2]Department of Mechanical Engineering, Faculty of Engineering, Kharazmi University, Tehran 1571914911, Iran
[3]Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC 3800, Australia

Corresponding author: Faezeh Marzbanrad (faezeh.marzbanrad@monash.edu)

ABSTRACT Major depressive disorder (MDD) has been considered a severe and common ailment with effects on functional frailty, while its clear manifestations are shrouded in mystery. Hence, manual detection of MDD is a challenging and subjective task. Although Electroencephalogram (EEG) signals have shown promise in aiding diagnosis, further enhancement is required to improve accuracy, clinical utility, and efficiency. This study focuses on the automated detection of MDD using EEG data and deep neural network architecture. For this aim, first, a customized InceptionTime model is recruited to detect MDD individuals via 19-channel raw EEG signals. Then a channel-selection strategy, which comprises three channel-selection steps, is conducted to omit redundant channels. The proposed method achieved 91.67% accuracy using the full set of channels and 87.5% after channel reduction. Our analysis shows that i) only the first minute of EEG recording is sufficient for MDD detection, ii) models based on EEG recorded in eyes-closed resting-state outperform eyes-open conditions, and iii) customizing the InceptionTime model can improve its efficiency for different assignments. The proposed method is able to help clinicians as an efficient, straightforward, and intelligent diagnostic tool for the objective detection of MDD.

INDEX TERMS Time series classification (TSC), major depressive disorder (MDD), EEG signal processing, deep learning, InceptionTime, diagnosis system.

I. INTRODUCTION

Mental health conditions have a significant influence on the quality of life from childhood to adolescence and adulthood [1]. Globally, especially the western countries, mental disorders are among the leading causes of disability, accounting for 30-40% of chronic sick leave and cost around 3% of gross domestic product (GDP) [2].

Major depressive disorder (MDD), which is also known as (unipolar) depression, is one of the most common mental health conditions [3]. It is characterized as a mood disorder that causes sleep disorders, interest deficit (anhedonia), energy deficit, feeling the emotion of sadness (low mood), poor appetite or overeating, changes in cognition, and vegetative symptoms for at least two weeks [4]. Based on these symptoms, particularly low mood and anhedonia as the two fundamental symptoms, a patient is diagnosed with depression [5]. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6], there are many side effects related to depression, such as significant weight loss or weight gain. Depression ranked first globally in Years lost to disability (YDL), while more than 300 million people suffer from depression across the globe [7]. It is also the ninth rank in disability and death together [8] since suicide is more common among individuals with depression [9]. These statistics have been exacerbated after the COVID-19 pandemic rapidly swept across the world in 2019. Studies, including the general population from five countries, have shown that the prevalence rate of depression has surged during the COVID-19 pandemic [10]. Also, it has been revealed

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam.

that the correlation between COVID-19 and depression for different groups is increasing, including medical staff with 18.4% self-reported rates of depression [11], and university students with 82.4% depression rate [12].

Generally, psychiatrists diagnose depression using either asking about individuals' symptoms, thoughts, feelings, and behavior patterns or by checking their questionnaires. The widely used criteria to diagnose depression are stated in DSM-5 [6], published by the American Psychiatric Association. However, this approach to diagnosis has two main limitations. First, clinicians and psychiatrists detect depression based on the information provided by the subjective claims of patients, which can cause the diagnosis to be biased. Second, the criteria determined by DSM-5 are insufficient in some cases or even heterogeneous in a number of diagnostic categories [13]. Various attempts have been frequently made to address the shortcomings of DSM-5 [14]–[17], using alternative approaches based on laboratory tests. For instance, there have been efforts to analyze metabolic changes in the body, such as urinary metabolites, to detect depression [18], [19]. Nonetheless, the most investigated alternative approach to diagnose depression is through analyzing brain activity.

Electroencephalogram (EEG) is a widely used method to record the electrophysiological dynamics of the brain, resulting from neural activity in real-time, by which the cognition, brain function or dysfunction and possible indication of mental disorders can be analyzed [20]. A typical EEG contains multiple sensors (electrodes) placed on different areas of the scalp [21]. These sensors pick up the activity of neurons in different parts of the brain, where the voltage variation between every two electrodes is recorded [22]. It has been proven helpful in the diagnosis of neurological, cognitive psychology, and psychophysiological disorders [23]. It can also be used in less developed or developing countries, where people are faced with a lack of physicians and psychiatrists. EEG recording can be automatic, adaptable to different contexts, portable, easy to follow, and cost-effective [24]. However, the interpretation of EEG data is not straightforward; due to the noise, variability between individuals, and substantial changes over time, even for the same person.

This study proposes a high-performance deep learning-based method to classify individuals with and without MDD using EEG signals. To develop the novel automated system, 30 healthy and 34 MDD individuals' EEG records were used. Afterward, we utilized the channel-selection method to reduce the required number of electrodes and find the most useful channels for the MDD recognition task. Besides, a sensitivity analysis was applied to investigate the effectiveness of the length and condition of EEG records as well as the model's parameters.

The main *contributions* of the proposed method can be summarized as follows:

A. Developing a generalized and robust intelligent system that accepts EEG time series data to detect MDD. To date, most EEG classifiers used high cost and memory usage processing, such as image-based and feature extraction-based methods, to enrich EEG records as the input of classification models. Therefore, there is a lack of straightforward approaches for implementation of MMD classification in real-world with timely response.
B. Proposing a re-organized and re-developed InceptionTime-based model for this application and comparing its performance with the incipient InceptionTime model. Not only is implementing our time series method innovative for diagnosing MDD, but this is also the first attempt to customize and analyze the InceptionTime architecture in a way that achieves superior performance in MDD detection.
C. Investigating an efficient EEG-based MDD detection, requiring a fewer number of electrodes by removing redundant channels. This is the first study, to the best of our knowledge, to consider whether reducing the number of electrodes is still promising in the performance of the model for MDD classification and in that case, which electrodes are better candidates to be kept. Indeed, a fewer number of electrodes is easier to implement, while being time and computationally cost-effective and less susceptible to noise.
D. Analyzing the effect of EEG recording conditions, segments, and the deep model structure on the classification performance. This is a leading attempt to argue whether shorter segments of EEG recordings can still hold the performance of the automated MDD classifier. It is essential to realize which segment of the EEG data conveys the most valuable information so that the model can only rely on that segment. Moreover, we investigated how different recording conditions, such as eyes being open or closed, affect the MDD classification model and whether physicians could only adhere to preferable conditions using the proposed model.

The abbreviations used throughout this paper are listed in Table 1.

II. BACKGROUND

In this section, some of the publicly available datasets are introduced, followed by reviewing relevant articles. Cavanagh *et al.* [25] presented an EEG recording dataset including 46 MDD individuals and 75 healthy controls. With a 64-electrode cap, EEG data were recorded from participants aged only between 18 to 25 years. A probabilistic task was assigned for the individuals when the EEG data were recorded, which is relatively energy- and time-consuming for participants and clinicians compared to resting-state recording data. Furthermore, Cai *et al.* [26] introduced an EEG dataset consisting of 24 MDD individuals and 29 healthy controls. There were 33 males and 20 females, and the subjects were aged from 16 to 56 years old. A 128-electrode cap was utilized to record EEG data only in eyes-closed condition during the resting-state. Wu *et al.* [27] collected

TABLE 1. Abbreviations list.

Abbreviation	Definition
ASR	Artifact Subspace Reconstruction
CNN	Convolutional Neural Network
DMN	Default Mode Network
DSM-5	Diagnostic and Statistical Manual of Mental Disorders
dDTF	direct Directed Transfer Function
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
EC	Eyes-Closed
EO	Eyes-Open
FDMB	Frequency-Dependent Multi-layer Brain
GPDC	Generalized Partial Directed Coherence
GAP	Global Average Pooling
GDP	Gross Domestic Product
HC	Healthy Control
HUSM	Hospital Universiti Sains Malaysia
kNN	k-Nearest Neighbors
LE	Linked Ear
MDD	Major Depressive Disorder
MAD	Mean Absolute Difference
NCA	Neighborhood Component Analysis
1DCNN-LSTM	One-Dimensional CNN-Long Short-Term Memory
PDC	Partial Directed Coherence
SSRI	Selective Serotonin Reuptake Inhibitor
STFT	Short-Time Fourier Transform
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TSC	Time Series Classification
WCOH	Wavelet Coherence
WT	Wavelet Transform
YDL	Years Lost to Disability

EEG recordings data using 32 channels. The number of subjects was higher than other databases, with 200 individuals for each MDD and healthy group. However, it seems not very generalizable as MDD individuals were aged 52.85 and 54.90 years old on average for women and men, and healthy ones were 49.87 and 54.59 years old, respectively. These age profiles of the participants are even higher than the middle ages of countries with the most geriatric populations [28]. That is to say, this dataset belongs to a distinct cohort, and thus it does not consider a universal one.

Mumtaz *et al.* [29] published a valuable public database with 19-channel EEG recordings of 34 patients with MDD and 30 healthy individuals with a proper distribution of gender, age, and the number of participants in different classes. The data includes both eyes-open and eyes-closed during the resting-state with a mean age of 40.3±12.9 and 38.3±15.6 years for MDD and healthy groups, respectively. They described the database and study design in their paper [29], where they also proposed a machine learning approach for MDD detection. They extracted a feature matrix using wavelet transform (WT) analysis and then reduced its dimension based on a rank-based feature selection method. The most effective features were given to a logistic regression classifier, which was trained through 10-fold cross-validation. The proposed method of this study achieved 87.5% accuracy, 95% sensitivity, and 80% specificity for the MDD classification task. Applying WT analysis may reduce the irrelevant information since it takes advantage of converting the data into compressed parameters. Yet, this approach is subjective due to the WT works on window functions, which are made of priory choice of frequency and time scales. Therefore, it requires experience in selecting predefined window functions and consequently depends on the analysis and EEG data.

The database published by Mumtaz *et al.* [29] was also used in several other research works. Mahato and Paul [30] developed machine learning models to classify MDD individuals with EEG signals, including support vector machine (SVM), logistic regression, Naive Bayes, and decision tree. Using features ranging from alpha (8-13 Hz), alpha1 (8-10.5 Hz), alpha2 (10.5-13 Hz), beta (13-32 Hz), and delta (0.5-4 Hz) to alpha band power with paired theta asymmetry, they found the highest classification accuracy was achieved using alpha2 power, compared to alpha and alpha1 power. Of all classifiers, SVM provided the best performance with an accuracy of 88.33%, where the combination of alpha2 power and theta symmetry were used as input features. Given that they used only eyes-closed records of 30 MDD and 30 healthy individuals, it seems their machine learning model's performance can be improved using more input data.

Aydemir *et al.* [31] also proposed a classification model based on hand-crafted features in three steps to detect MDD. They used melamine patterns and discrete wavelet transform (DWT) to generate features. The most valuable features were selected using neighborhood component analysis (NCA). Finally, Quadratic SVM and weighted k-nearest neighbors (kNN) were able to classify individuals with a 99.11% and 99.05% accuracy as their superior results. While using DWT results depend on the number of decomposition levels, the more the decomposition levels are, the more computationally complex and time-consuming the model is. Additionally, despite the melamine pattern being a novel idea and having good results, it should be further assessed to confirm its performance.

Dand *et al.* [32] used a frequency-dependent multi-layer brain (FDMB) network along with a convolutional neural network (CNN) to determine if an individual has MDD. The time-frequency characteristics were extracted from EEG signals, where each frequency band corresponds to a single layer of a multi-layer network. The FDMB network takes advantage of the frequency characteristics and channel coupling of EEG signals simultaneously to provide the input of a CNN-based architecture. In terms of input data, they first selected 0-180s of records and then used a non-overlapping sliding window with a length of 2s to segment the samples. They achieved an accuracy of 97.27% and specificity of 97.33% for MDD detection. Despite using three convolutional layers, they showed that there was less than a 0.5% difference in performance with one or two convolutional layers in the core block.

Saeedi *et al.* [33] introduced five different deep learning architectures to discriminate MDD individuals from healthy controls. They extracted the association between EEG channels using generalized partial directed coherence (GPDC) and direct directed transfer function (dDTF) methods in the form of effective brain connectivity analysis. Also, a new approach for the combination of sixteen connectivity methods (GPDC and dDTF in eight frequency bands) was utilized to construct an image for every individual. The constructed images of eyes-closed records were used to train and test their models. The experimental analysis shows that a one-dimensional CNN-long short-term memory (1DCNN-LSTM) model outperforms all models, with an accuracy of 99.245%, a sensitivity and specificity of 98.519%, and 100%, respectively. Although their evaluation of the 2DCNN-LSTM method led to a faster model, it was less efficient than 1DCNN-LSTM, even with more parameters. Therefore, it is worthwhile to investigate a top-performing model by considering both the performance and computation time of the models.

Loh *et al.* [34] suggested a CNN model for detecting MDD patients from healthy individuals using an imaging approach. At the first step, spectrogram images were obtained after STFT was applied to EEG signals, resulting in 3,600 images from the individuals in the dataset. These images were then passed through an eight-layer CNN network to detect MDD. Their approach achieved 99.25% accuracy, 99.24% sensitivity, and 99.26% specificity after utilizing 10-fold cross-validation. Their model achieved high performance, yet might not be practical to implement in clinics, since the 2D-CNN-based models have high computational costs.

Khan *et al.* [35] estimated the effective connectivity within the brain default mode network (DMN) regions using a partial directed coherence (PDC) method to properly classify MDD individuals. They first used the continuous 2-second window of records and calculated the PDC matrices based on them. Next, they discarded all indirect causal effects of non-DMN regions by extracting the connectivities. Finally, a three-dimensional (3D) CNN took in the PDC matrices as its input for the classification task. They performed classification based on the number of labeled PDCs. Using a 10-fold cross-validation technique, they attained an accuracy of $94.96 \pm 7.32\%$ for the proposed classification algorithm. In the next work, Khan *et al.* [36] constructed a 2D-CNN-based deep learning model with the ability to distinguish between MDD and healthy individuals. They selected DMN channels and extracted continuous 2-second segments from 19 channel pre-processed EEG data. Afterward, they applied wavelet coherence (WCOH), to determine the coherence of these segments, mapped them to the format of image for each individual. These images were the input of the 2D-CNN architecture, which consists of 5 convolutional layers with a kernel size of 5×5 for all layers. To evaluate their model, a 10-fold cross-validation technique was considered, which led to an accuracy of 96.0 ± 1.55. Using novel feature extraction methods, PDC in the first paper, and WCOH in the second paper, yield remarkable results. Nonetheless, for extending these outcomes, more assessments should be applied to verify their model performance.

To date, a variety of models have been developed to detect MDD involving CNN-based and non-CNN-based ones. Although CNN-based models showed promising performance, they generally used image conversion, sub-sampling, and complex mathematical calculations to provide the input data for their models. Non-CNN-based models, however, applied traditional machine learning models with a focus on the extraction of handcrafted features. Even though they tried to select the most crucial features and mostly considered a more extended segment of records, their models' performance was not very promising and relied on all channel data. Therefore, we have tried to bridge this gap by introducing a model that can directly work with EEG time series to dramatically decrease the computational costs and time for automatically detecting MDD with a notable performance. This approach offers the flexibility of investigating the condition and length of EEG recording. Therefore, we have re-organized a time series model for MDD classification, which can operate without any pre-extracted features. Besides, as a different number of electrodes and caps are being used to record EEG signals, it is crucial to assess which electrodes have a more decisive influence on MDD classification. Hence, we propose a three-step channel-selection approach to discover the most effective electrodes and compare the performance of the model using this data. To summarize, these novel attempts help achieve lower computational cost and memory usage and a more straightforward and convenient data collection process for the patients and physicians.

III. MATERIALS AND METHODS

A. DATASET

In this study, we have used a publicly available dataset provided by Mumtaz *et al.* [29] in 2017. The dataset comprises multi-channel EEG recordings obtained from 64 participants, including 34 MDD individuals (17 females and 17 males with a mean age of 40.3 ± 12.9 years) and 30 healthy control (HC) individuals (9 females and 21 males with a mean age of 38.3 ± 15.6 years). All MDD subjects were labeled according to the DSM-5 criteria [6]. Participants were informed about the experiment in advance, and completed written consent forms. Moreover, the human ethics committee of Hospital Universiti Sains Malaysia (HUSM) approved the experiment procedure. In order to reduce the effect of medications on the recorded data, the MDD individuals did not receive any medications for two weeks before the experiment, yet given antidepressants under the general category administered by Selective Serotonin Reuptake Inhibitors (SSRIs), along with the consultation of a psychiatrist. A nineteen electro-gel sensors EEG cap was applied to acquire EEG data according to the 10-20 electrode placement system [21], referencing the linked ear (LE). In this regard, the on-scalp placements of electrodes, which can be subdivided into separate regions, including the frontal (Fp1, Fp2, Fz, F3, F4, F7, F8), temporal (T3, T4, T5, T6), parietal (P3, P4, Pz), occipital (O1, O2) and

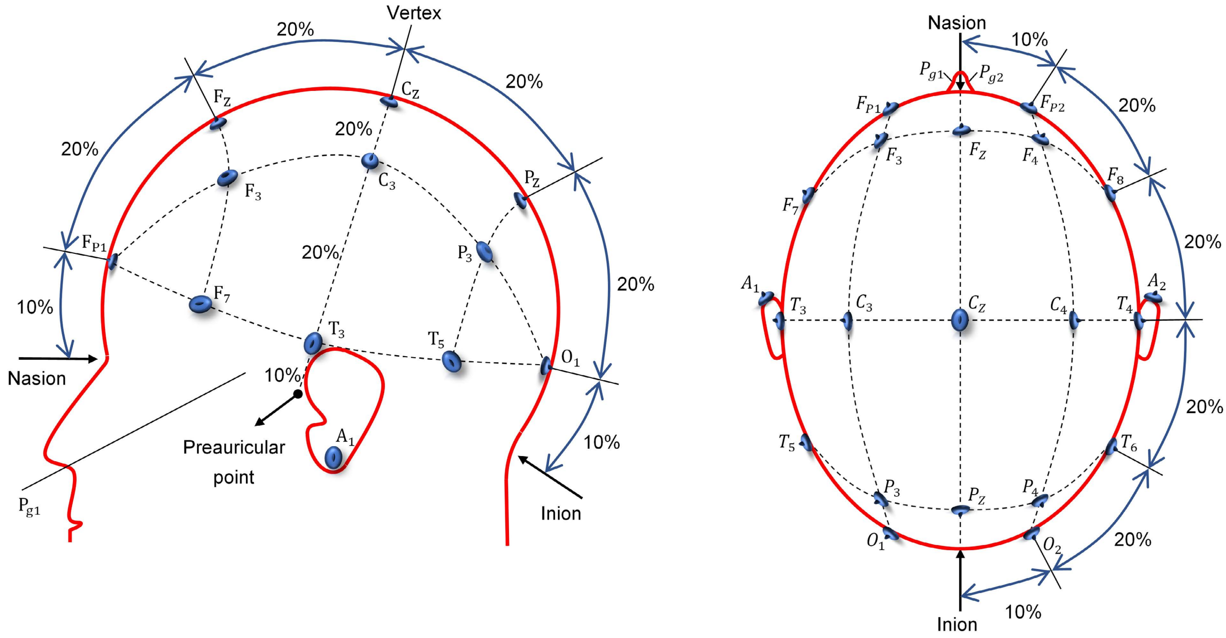

FIGURE 1. The schematic of electrodes location according to the 10-20 international system.

central (C3, C4, Cz) areas, and reference points electrodes (A1, A2) are represented in Fig. 1. The sampling frequency of the EEG data was set to 256 Hz.

In the experiment, the EEG data of 5-min-long eyes-open (EO), 5-min-long eyes-closed (EC), and ten-min-long TASK conditions of each individual were collected. The participants were instructed to keep their eyes open and closed with minimal blinking and head movement for five minutes during EO and EC, respectively. During the TASK condition, the participants were asked to fulfill a ten minutes task of visual stimulus, where they entered the 'SPACE' button on the keyboard each time the target flashed on the screen. The EO and EC data are not available for every individual, but each participant has at least one of them. We only focused on EO and EC records for this study to train the proposed model for diagnosing MDD individuals to see if MDD could be detected without performing the TASK test.

B. PRE-PROCESSING

Raw EEG data are usually contaminated by noise and artifacts. Thus, several filters were applied to avoid subsequent errors in analysis, ensuring that the underlying neuronal activity is truly represented by the data. At the first step, a bandpass filter with cutoff frequencies of 0.1 Hz and 70 Hz was applied to the data, followed by a notch filter to suppress 50 Hz power line noise. Therefore, the frequency bands which do not have considerable influence on the MDD classification, and repetitive spectral noise were removed [37]. Detecting and correcting the background artifacts due to eye blinks, eye movements (vertical and horizontal), muscle and heart activities, and sensor motions were the next step. For this purpose, the artifact removal algorithm was implemented based on the artifact subspace reconstruction (ASR) method [38]. ASR is an automatic, adaptive method that can effectively wipe large-amplitude or transient artifacts comprising multichannel EEG recordings. Furthermore, since some of the available records have more than nineteen rows of data, we discarded excessive channels. These extra channels of data were recorded either as additional data or as a reference, such as A1 or A2 channels, which play the role of reference electrodes by locating at the ears. After processing the data, we selected the first four minutes of each EEG record and omitted records less than four minutes long. This is mainly because there is not usually an exact five-minute recording in the dataset, and most of the records consist of less than five minutes long. Also, we wanted to use as long as possible input data to develop our model. Afterward, we mapped data of each record between −1000 and 1000 mV. Amplification of the EEG data could help the deep learning model to discover clues for the identification of MDD individuals. All these steps have been done using MATLAB software (version 2018a) and EEGLAB built-in plugin [39]. The overall pre-processing pipeline of our work is presented in Fig. 2, which provides 115 pre-processed EEG records.

C. CLASSIFICATION

Since EEG classification-based problems, as a time series classification (TSC), differ substantially from traditional supervised learning models for structured data, the temporal information of the signal should be considered. A novel and state-of-the-art deep learning ensemble for TSC is InceptionTime [40], which is the equivalent of the AlexNet [41] model for TSC. It has been shown this model potentially has a promising performance for multidimensional time-series data with different length sizes [42]. InceptionTime is an ensemble of five distinct TSC deep learning models, each of

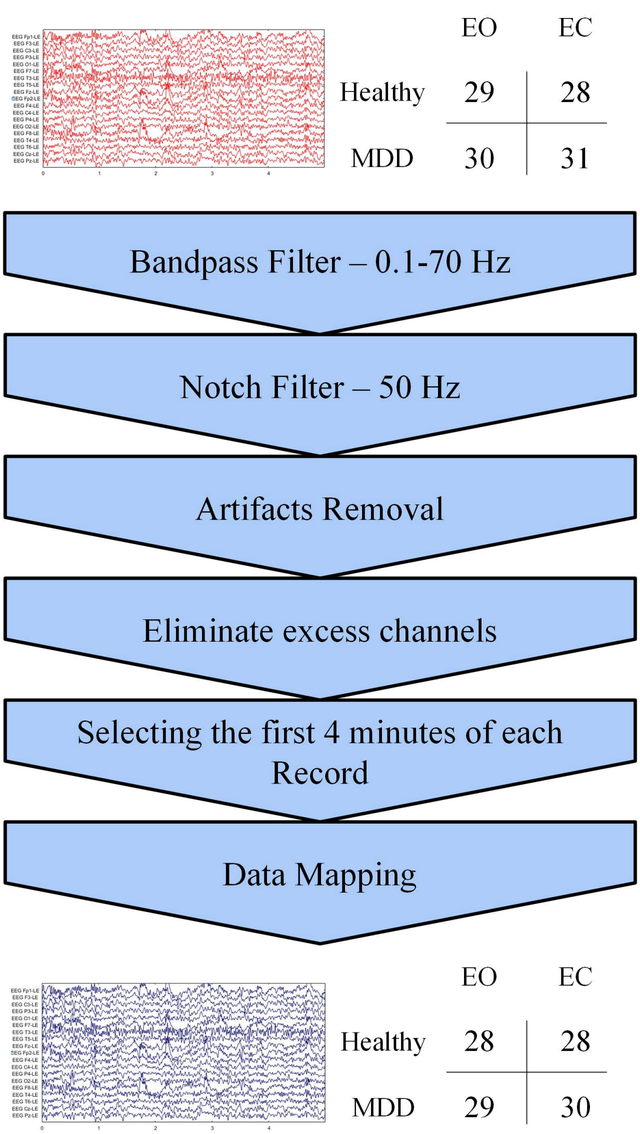

FIGURE 2. **The schematic of electrodes location according to the 10-20 international system.**

which is generated by cascading multiple Inception modules. The basic idea of the Inception module is to utilize various filters to a time series input simultaneously. This module includes filters in varying lengths and allows the network to extract relevant patterns automatically from both short and long time series. In this research, we have borrowed the idea of the InceptionTime model and implemented a modified and customized version of it to distinguish between MDD and healthy individuals using EEG signals.

The used Inception network classifier comprises two different residual blocks, each of them consisting of three Inception modules instead of traditional fully convolutional layers. The input of each residual block is transferred to be added to the next block's input via a shortcut linear connection, which leads to mitigating the gradient vanishing problem by occurring a direct gradient flow [43]. These residual blocks are followed by a global average pooling (GAP) layer that averages the multivariate time series' output over the whole-time dimension. At the end of this architecture, just a direct fully-connected layer with two neurons representing Healthy and MDD classes is used with the softmax activation function.

Fig. 3 illustrates an architecture of the Inception network with six separate Inception modules stacked one after the other. Also, it comprises the inside details of the proposed Inception module operations. The bottleneck layer serves as an operation of sliding filters to transform the dimension of the EEG time series in the Inception module. Thus, this not only can reduce the complexity of the model and mitigate overfitting problems, but it also provides much longer filters for the Inception network with roughly the same number of learning parameters in comparison with other methods. We have considered a bottleneck layer for each Inception module with a size of 57. The modification of the bottleneck layer from 1 to 57 led to a tremendous increase in each Inception module's dimension. Therefore, a higher number of sliding multiple filters with different lengths have been applied simultaneously on the same input time series. In this regard, we have utilized three sets of 64 filters of lengths 10, 20, and 40. Additionally, a parallel MaxPooling operation with a considered size of 3, followed by a bottleneck layer, has been applied to reduce the dimensionality. The output of the MaxPooling window has been calculated by taking the maximum value in the given window. Afterward, the output of each independent layer has been concatenated to form the output of the multivariate time series. The same operations have been repeated for every Inception module of the network. Eventually, five Inception networks formed the proposed deep model. This is mainly because an Inception network shows a high standard deviation in the accuracy due to the stochastic optimization process and randomly initialized weights during the training process. Moreover, we have considered the mean squared error as the loss function of the proposed model instead of the categorical cross-entropy of the InceptionTime and the stochastic gradient descent (SGD) by applying the momentum and the Nesterov techniques rather than the Adam optimizer for the training phase [44], [45]. The evaluation of the proposed model's parameters has been done in Section V.A (Table 6 and Table 8).

D. EEG CHANNEL-SELECTION

After attaining a robust and high-performance deep learning model to classify individuals with 19 channels, we aim to explore the possibility of classification with a smaller number of channels (electrodes). Feature extraction techniques were used to discover the most effective channels for discriminating between two classes. These methods identify the features, which seem to be less relevant for the objective, irrespective of the utilization of any learning algorithm. They are relatively fast and inexpensive in terms of computation and are helpful for recognizing correlated, duplicated, and redundant features. However, these methods cannot detect multicollinearity, especially when a combination of features

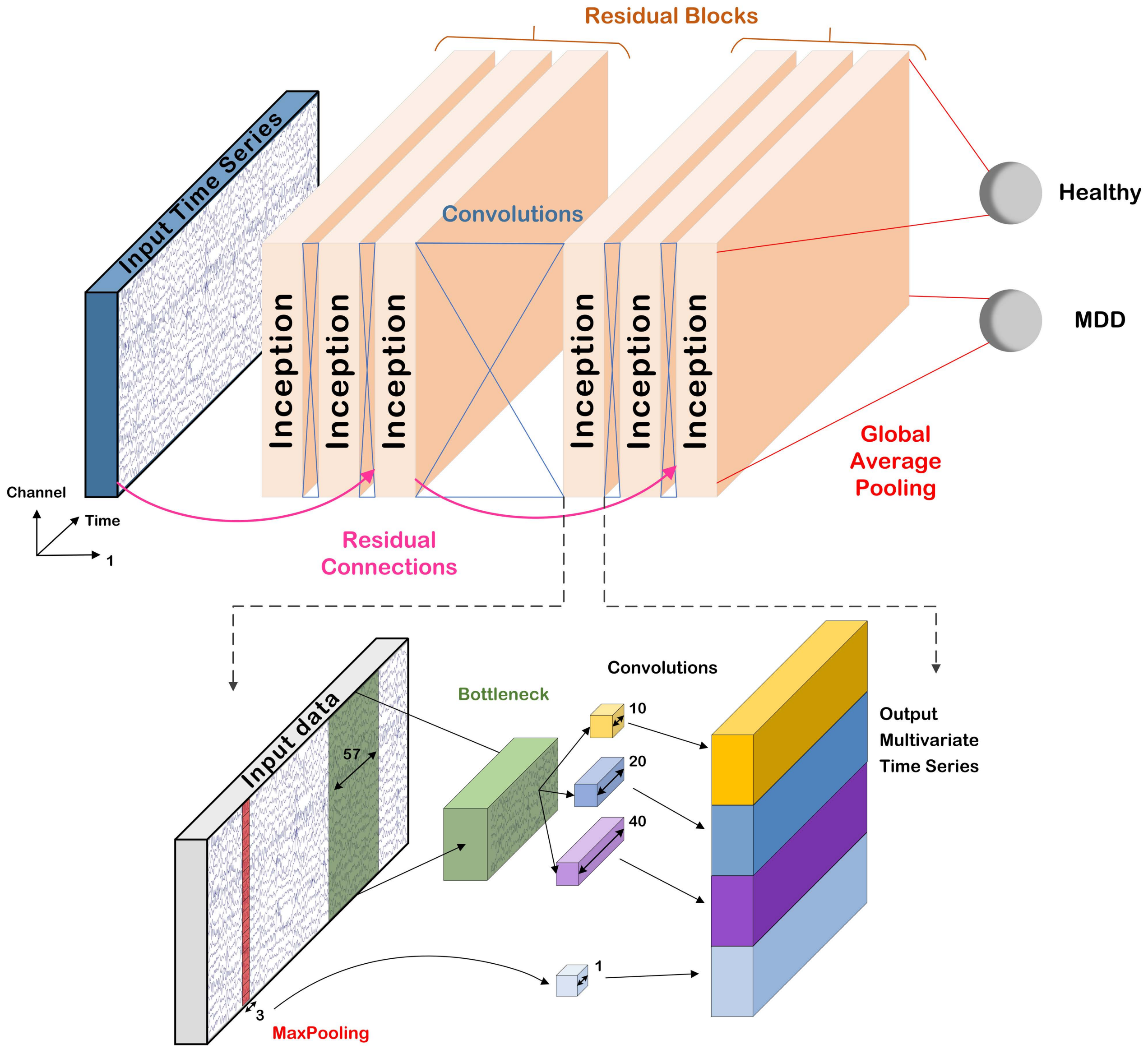

FIGURE 3. The proposed Inception network architecture with inside details of each Inception module for the EEG time series classification task.

may yield an enhancement in the overall model performance. So, we have applied three complementary feature selection steps to find the most effective channels, as described in the following sections.

1) MEAN ABSOLUTE DIFFERENCE

We have calculated the mean absolute difference (MAD) for each channel of all participants' EEG signals at the first step of the feature selection [46]. This approach calculates the absolute difference of a channel data from the mean value, which is defined as (1):

$$\mathrm{MAD}_i = \frac{1}{n}\sum\nolimits_{j=1}^{n}\left|X_{\mathrm{ij}} - \bar{X}_i\right| \tag{1}$$

where X_{ij} is the jth signal value of the ith channel, $\bar{X}_i$ is the mean of values of the ith channel, and n is the number of signal values. The major distinction between the MAD and variance threshold measures is the lack of the square in the former. Fig. 4 demonstrates the MAD value for each channel. The higher the MAD, the higher the discriminatory power of a

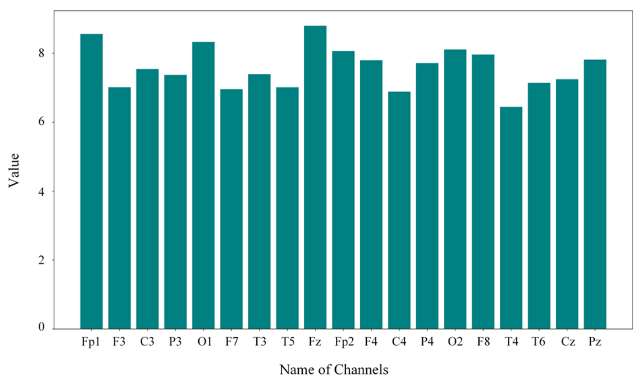

FIGURE 4. The MAD measure of each EEG channel.

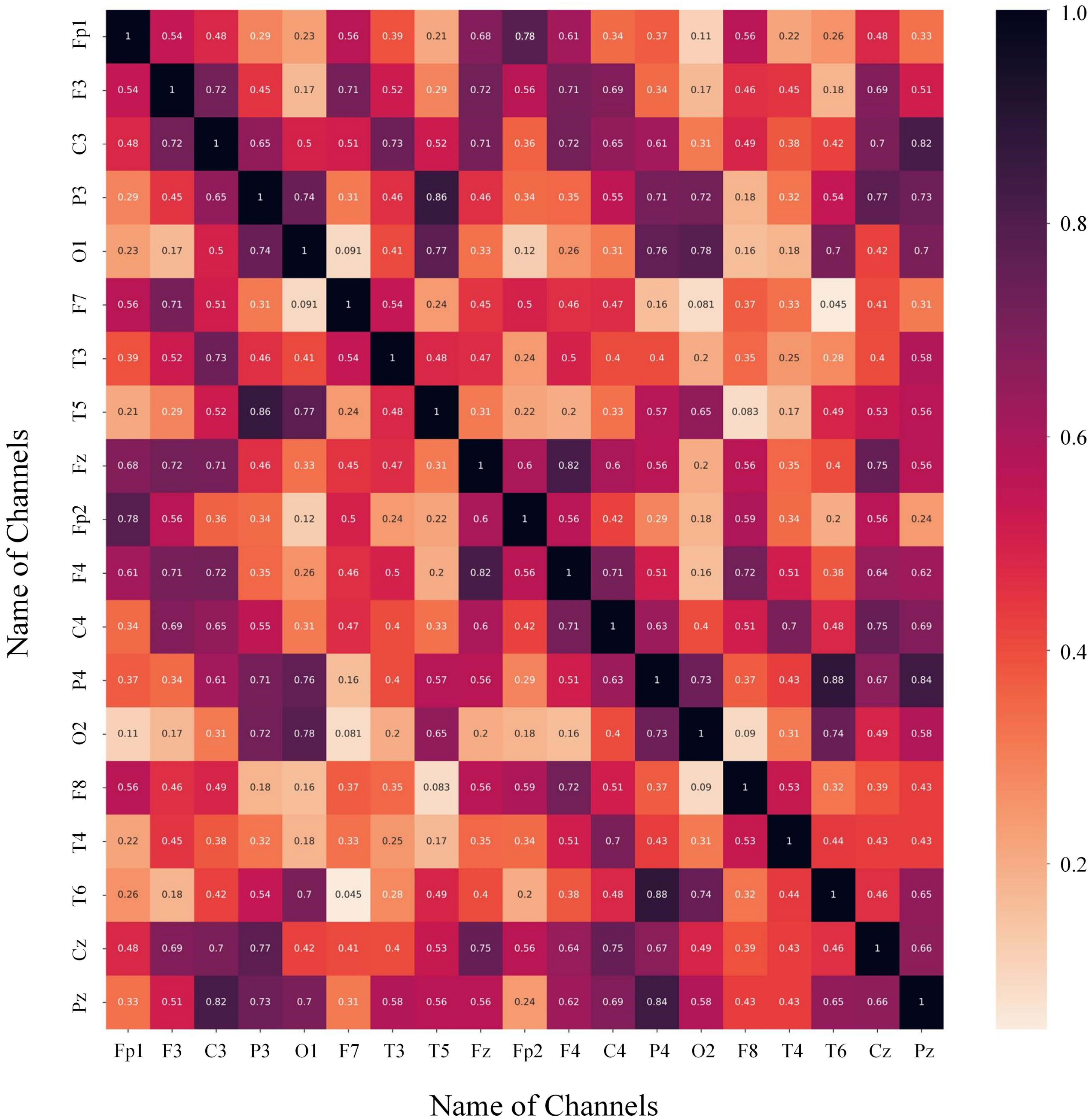

FIGURE 5. **The correlation coefficients matrix of channels.**

channel and potentially the more effective channel. As can be clearly seen in Fig. 4, since there is not a significant difference between the MAD values of the channels, it is hard to decide which channel should be omitted for the classification task. Nonetheless, the MAD values will be combined with the following experiment to select the most influential set of channels.

2) CORRELATION COEFFICIENT

In the second step, we have computed the correlation coefficients for different channels. A correlation-based algorithm represents a measure of the relationship of features, in which one feature can be predicted from the other, if they are correlated [47]. That is, the correlation coefficient demonstrates how a feature moves in relation to another. The higher correlation between two features, the more similarity between the features. Therefore, channels should be uncorrelated among themselves; otherwise, the learning model might only need one of those, as the second one is more likely to not accumulate additional information. Fig. 5 depicts the correlation matrix of all the EEG channels, which consists of a correlation coefficient between them. This value is computed as (2):

$$\rho_{xy} = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \tag{2}$$

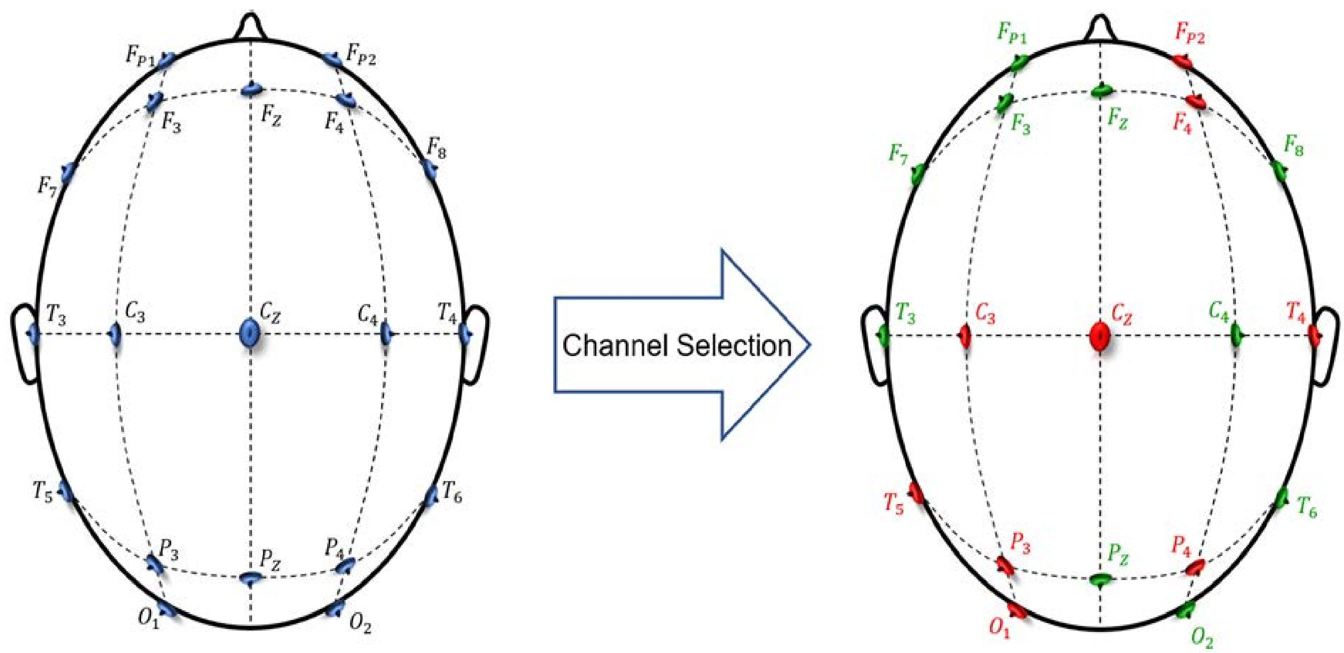

FIGURE 6. EEG channel-selection result. The red electrods represent omitted channels.

where ρ_{xy} represents the correlation coefficient between x and y channels, σ_x is the standard deviation of channel x, and σ_y is the standard deviation of channel y. To explore correlated channels, we have set a correlation coefficient of 0.75 as the threshold for dropping one of the two correlated channels. Table 2 shows channels with their correlated pairs.

First, we have omitted a channel that correlates with three or more channels. Regarding this matter, O1, P4, and Cz channels have been discarded. Next, if the correlation coefficient of two channels is equal to or greater than 0.75, then the MAD measure determines which of them should be excluded. At this stage, we have omitted the channel with the lower MAD measure. Thus, Fp2, C3, T5, and F4 channels have been discarded in the step of the correlation coefficient channel-selection approach.

TABLE 2. Name of channels and their correlated pairs.

Channels	Correlated pair(s)
Fp1	Fp2
C3	Pz
P3	T5, Cz
O1	P4, O2, T5
T5	P3, O1
Fz	F4, Cz
Fp2	Fp1
F4	Fz
C4	Cz
P4	T6, Pz, O1
O2	O1
T6	P4
Cz	P3, Fz, C4
Pz	C3, P4

3) BACKWARD-ELIMINATION ALGORITHM

To reach the aim of an appropriate classification with the lowest number of channels, we have employed the backward-elimination algorithm [48] to obtain the ten most influential channels. In comparison with two previous steps, which use the intrinsic properties of the channels via univariate statistics, this method explores all possible subsets of channels, assessing their performance by training and evaluating the classifier using that subset. This algorithm follows a greedy search approach by testing all possible combinations of channels by removing one channel at a time, and comparing the accuracy of the models when they are trained. Therefore, channel selection and classification have been performed for every individual subset and that with the lowest lost function has been selected. Each subset, at first, comprises 11 out of 12 channels to find the most inefficient channel. We have continued this process after removing one channel with 10 out of 11 remaining channels to eliminate another. At the end of this process, T4 and P3 channels have been excluded. Eventually, Fp1, F3, C4, F7, T3, Fz, O2, F8, T6, and Pz channels have been kept for training and evaluating the model as ten EEG channels to classify MDD individuals. Fig. 6 demonstrates the locations of omitted electrodes on the scalp in the 10-20 international system.

IV. RESULTS

The evaluation of the proposed method comprises two different experiments. First, we have randomly split the data set into the train (24/115), validation (14/115), and test (77/115) sets. This segmentation is the same for the following evaluations in which the test set consists of the same proportion of data types, 6 EEG records for each EO and EC of the two labels. In the training phase, a shuffled batch

TABLE 3. Hyper-parameters of the customized inceptionTime model for the MDD classification Task.

Hyper-parameter	Amount
Depth	6
Kernel size	40
Bottleneck	57
Number of filters	64
Maxpooling	3
Batch size	32
Learning rate	0.01
Momentum	0.9

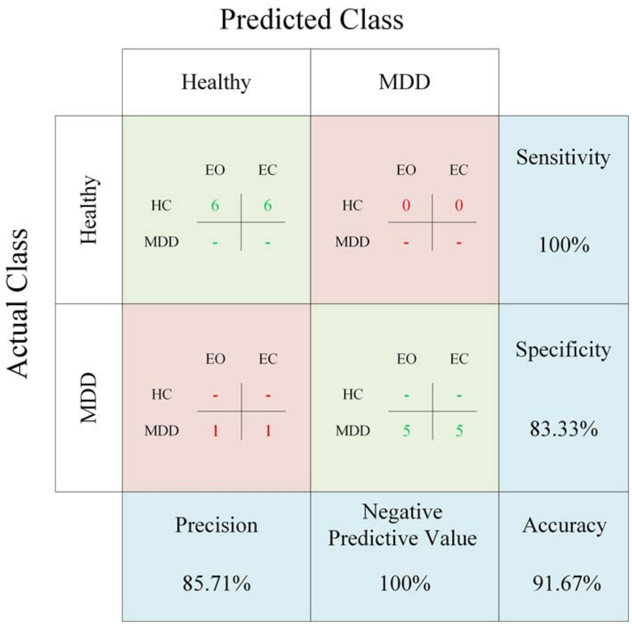

FIGURE 7. Performance metrics of the model in the face of 10-channel EEG data.

size of 32 updates the network weight through minibatch SGD using the learning rate of 0.01 with the momentum of 0.9 and the Nesterov technique. Table 3 delineates the details of the model's hyper-parameters. The model has been implemented for all experiments in Python using Keras framework 2.7.0 with Tensorflow 2.7.0 as the backend on the Google Colaboratory.

Subsequently, the model was trained with the 19-channel EEG data. Fig. 7 illustrates the confusion matrix, sensitivity, specificity, accuracy, negative predictive value, and precision of the test data of the trained model with 19-channel EEG data. As clearly can be seen, the model was able to classify all healthy records precisely, whereas it was unable to recognize two MDD records, one EO and one EC recordings. That is, the model demonstrated the highest possible sensitivity in the face of the test data; however, two false detections of healthy class as MDD caused the specificity of 83.33%. The accuracy metric stood at just under 92%, and negative predictive value and precision indicated 100% and 85.71%, respectively.

Moreover, the evaluation of the channel-selection approach was considered in the first experiment. The model performance for classifying MDD individuals using ten selected channels is represented in Fig. 8. This model misclassified

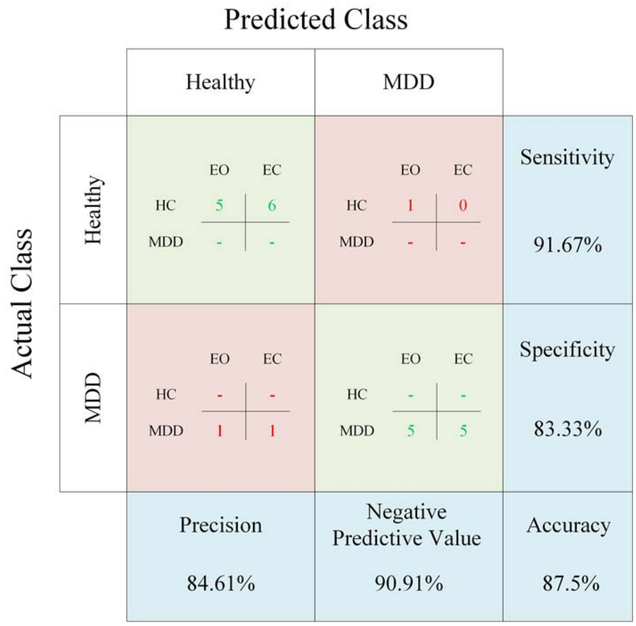

FIGURE 8. Performance metrics of the model in the face of 19-channel EEG data.

one EO and one EC recordings for MDD, and an EO recording for a healthy individual. The sensitivity and specificity of the model are 91.67% and 83.33%, respectively. Using 10-channel EEG data, the developed model can discriminate MDD's records from healthy ones with an accuracy of 87.5% and a precision of 84.61%.

In our second experiment, to have a more concrete illustration and a better comparison, we have employed a stratified 10-fold cross-validation at the subject level and have chosen two state-of-the-art research for comparison [29], [30]. Individual-wise partitioning (involving both EO and EC of an individual in either train or test sets) prevents the possible occurrence of being one record of an individual in the train set and the other record in the test set. In addition, we have selected the two proposed models due to several contributing factors. They have utilized the same dataset as ours, considered the EEG time series data (2D data) instead of short-time Fourier transform (STFT) image processing and other imaging approaches without using sub-samples, and tried to use feature selection to classify with the lowest possible features. Thus, we have calculated the accuracy, sensitivity, and specificity of our proposed methods before and after channel selection and compared them with their results in Table 4.

V. DISCUSSION

The utilization of artificial intelligence-based methods has been proliferating in healthcare problems [49]. More specifically, since the number and quality of EEG records are increasing, the application of these methods has become popular. We proposed a deep learning approach for automated detection of MDD individuals using 19-channels EEG data.

TABLE 4. Classification results of our proposed methods using stratified 10-fold cross-validation and two state-of-the-art research on the same dataset and similar input type.

Models	Features	Accuracy	Sensitivity	Specificity
Mumtaz et al. [29]	Wavelet features	87.5% (±7.1)	95% (±4.3)	80% (±8.8)
Mahato et al. [30]	Alpha 2 + Paired Theta asymmetry	88.33%	89.41%	90.81%
Our proposed method	19-channel EEG data	91% (±4.9)	94.9% (±8.2)	88.2% (±11.3)
Our proposed method after channel selection	10 selected channels EEG data	90.1% (±5.5)	93.2% (±8.8)	88.2% (±11.3)

Most of the studies on this topic have followed either feature extraction, time-frequency representation, or subsampling approaches. Research works based on feature extraction methods focused mainly on using band powers as well as other handcrafted features for the classification task using EEG data and machine learning techniques [29]–[31]. Other research works, which employ the STFT approach, were interested in applying image processing techniques on transformed EEG to image data. They usually involve a CNN-based architecture on STFT images to find the effective patterns for MDD classification [32]–[34]. The last group extracted EEG segment length (e.g., two-second segment) and developed a model to classify every segment. When all segments were classified, it would be possible to categorize the EEG data [35]. We considered MDD classification as a TSC problem and tried to find a solution using the EEG time series data without any conversion and extraction. Therefore, this approach can reduce the computation cost and time, occupy lower memory and make the classification task closer to real-time implementation. More specifically, as the proposed methods do not require particular algorithms and methods to prepare data for the classification task, it significantly decreases the resources necessary to run the model. In terms of performance, our model showed superior proficiency, especially with the sensitivity metric.

We compared our method with two previous techniques, which involved feature extraction but considered data as time series [29], [30]. Mumtaz *et al.* [29] compared three time-frequency decomposition analyses where they found out wavelet transform was the most suitable for feature extraction. According to their experiment, the most effective features were EEG data of frontal and temporal areas with delta and theta bands. This was consistent with our channel-selection method, which also found that EEG frontal channels were the most relevant (Fig. 6). They selected the first two minutes of both EC and EO EEG records to train and test their model. Mahato *et al.* [30] applied different types of features and separate machine learning techniques as the classifier for automated detection of MDD. The highest performance was achieved with SVM using a combination of alpha2 and theta asymmetry, which was trained and evaluated using EC EEG records. Our proposed method was able to use four-minute-long 19-channel EEG data as an input to classify MDD individuals with an accuracy of 91.67% on the test set. It can detect all healthy records accurately (sensitivity of 100%), but one EO MDD and one EC MDD were misclassified (Fig. 7). After subject-wise partitioning and 10-fold cross-validation, the model showed an accuracy of 91% and sensitivity of 94.9% (Table 3). Besides, it showed that the proposed deep learning structure of this study, can be implemented on EEG data with superior performance as a TSC model. Our proposed model, as a costumized version of the InceptionTime model, is different from the original model in a number of ways, including the hyper-parameters, structure of every single Inception module, and the training scheme. We also realized that the bottleneck hyperparameter of the model has the most effective parameters to achieve the best performance.

Classifying with the smaller subset of channels is usually a valuable approach as it might provide low computation cost and memory usage, reduced time, and better generalization capability. It is also beneficial for clinicians and patients as being more convenient in clinical practice. For automated detection of MDD individuals, we tried to minimize the number of channels. Thus, three distinct channel selection methods were applied to the EEG data. As well as the benefits of classification with the lower number of features, it still works if we combine with 64-channel EEG datasets and just use the selected channels for MDD classification. This approach can solve the problem of the lack of data for training deep learning models and their generalizability. As Fig. 6 shows the result of our work for finding the effective channels, there are no particular patterns in terms of the area where electrodes are located. Despite this point, the frontal area seems to have the most impact, and temporal and parietal areas possess the inferior impact on the classification of MDD individuals. Interestingly, neurophysiological studies also have shown the prefrontal cortex, which is a part of the brain's frontal lobe, often involves activity changes in individuals with depression and anxiety [50], [51]. In addition, the performance metrics of 10-channel EEG data experiments present a promising outcome. Although it has about 5% lower accuracy than the 19-channel experiment on the test set, there was less than a 1% difference between the two experiments in a 10-fold cross-validation approach. In the former investigation, the model misclassified one EO healthy individual through 10-channel EEG data compared to 19-channel EEG data, which leads to a lower specificity and negative predictive value (Fig. 8). Interestingly, as we can see in Table 2, both input data possess the same specificity, while 10-channel EEG data has a lower 1.7% sensitivity.

A. SENSITIVITY ANALYSIS

EEG data collection with relatively long records (e.g., five-minute records), is time-consuming and can add an increment of computation costs and memory usage for the recording and processing. Therefore, reducing the duration of the measurement not only mitigates these problems but also provides the opportunity of collecting more data. Hence, we have designed a sensitivity analysis to study the influence of different parts of recordings on the classification of MDD individuals and find the segments, which convey the most precious information. For this aim, we have cut the first, second, third, and fourth minutes of the EEG data. With the same model architecture and parameters, we have input these segments of data into the model separately and measured the model's accuracy, sensitivity, and specificity for the classification task. Table 5 shows the mean and standard deviation of each segment using a stratified 10-fold cross-validation technique for 19-channel EEG data.

TABLE 5. Classification results of our proposed method using the first, second, third, and fourth minutes of the EEG data.

Segments	Accuracy	Sensitivity	Specificity
First (0 – 1 min)	90.1% (±5.5)	96.6% (±10.5)	86.5% (±7.2)
Second (1 – 2 min)	88.4% (±7.2)	94.9% (±8.2)	84.9% (±12.2)
Third (2 – 3 min)	87.5% (±7.3)	93.2% (±8.8)	84.9% (±12.2)
Fourth (3 – 4 min)	85.7% (±7)	91.5% (±8.9)	83.3% (±13.5)

The outcomes demonstrate the fact that each segment of the time window comprises relatively adequate patterns and regularities to distinguish MDD individuals from healthy ones. Notably, the first minute of EEG data might convey the most precious information as the performance metrics of this time window are higher than other segments. Despite having a 0.9% accuracy lower than the whole record experiment, the first segment reveals a 1.7% higher average sensitivity. On the contrary, the fourth minute of the records had the lowest performance in the classification task. Interestingly, as we get closer to the last segments, the accuracy decreases. Besides, all segments show a sensitivity higher than 90%, which means they provide an accurate classification for healthy records.

There are two important results from this analysis. First, MDD classification can be fairly accurate by using every one-minute segment of data throughout the four-minute record instead of the whole record. Second, the first minute might convey the most precious information since the performance metrics of the model in this experiment are as desirable as the model's performance in a four-minute-long record. As a result, recording only one minute of EEG data might be enough for the MDD classification problem. This potentially addresses the issues mentioned earlier regarding collecting data and can provide the opportunity of gathering a higher number of individuals' EEG records with less inconvenience for the patients and clinicians. With a larger database, potentially more accurate, reliable, and generalizable machine learning models could be developed.

In a further experiment, we have investigated the depth parameter of the inception networks in our InceptionTime model using 19-channels EEG data. In this regard, we have kept unchanged all training parameters and only substituted the depth parameter. It represents the number of Inception modules in the Inception network, which is changed from 6 to 5 and 7 to analyze the impact on the model's performance. Table 6 illustrates the model performance using different depths with consideration of a stratified 10-fold cross-validation.

TABLE 6. Classification performance of the model with the different depth sizes.

Depth size	Accuracy	Sensitivity	Specificity
5	87.4% (±7.4)	91.5% (±8.9)	84.9% (±12.2)
6	91% (±4.9)	94.9% (±8.2)	88.2% (±11.3)
7	90.1% (±5.5)	93.2% (±8.2)	88.1% (±8.2)

Setting depth as 5 resulted in a 4.6% accuracy decline in comparison with the depth size of 6 and reached the sensitivity and specificity of 91.5% and 84.9%, respectively. While the accuracy and sensitivity of the model were similar in the depth of 7 to the original model with less than 1% decrease, the sensitivity stood at 93.2% with a more than 1% reduction. The most critical point when we enhance the depth is the increase of the computation cost and time considerations. Thus, a trade-off between the performance and computational cost is necessary to consider in terms of enhancing the depth. The depth of 6 for the model had both advantages of the performance and computational cost for the detection of individuals with MDD.

The third experiment was dedicated to examining the performance of the model according to the state at which EEG data was recorded. Thus, we first left out the EC records, evaluated the model's performance using only EO data, and then repeated the experiment using only EC. Table 7 demonstrates the performance metrics of the models, which were trained and tested using only EO, only EC, and the combination of them and stratified 10-fold cross-validation with the same architecture of the proposed model. The model with only EC records shows superior accuracy and specificity with 92.7% and 91.5%, respectively. Nonetheless, evaluating using EO records leads to lower performance, where the accuracy stands at 89%. Although the combination of both records seems to yield a more reliable model as it sees the higher number of records, EC recordings might be sufficient for the classification of MDD individuals.

Comparing the performance metrics of the original InceptionTime structure and parameters with our customized model deciphers how the proposed approach can be helpful. In this regard, we have trained and tested the InceptionTime model using a stratified 10-fold cross-validation technique and reported its accuracy, sensitivity, and specificity in Table 8. As clearly can be seen, our proposed methods outperform in all the performance metrics. Of note, it has a

TABLE 7. Classification performance of the model using different conditions of recorded EEG data.

Input EGG data	Accuracy	Sensitivity	Specificity
EO	89% (±7.7)	93.2% (±8.8)	86.6% (±13.1)
EC	92.7% (±4.8)	94.9% (±8.2)	91.5% (±8.9)
EO + EC	91% (±4.9)	94.9% (±8.2)	88.2% (±11.3)

TABLE 8. Classification performance of the inceptiontime model and our proposed method using 19-channel EEG data.

Input EGG data	Accuracy	Sensitivity	Specificity
InceptionTime	79.6% (±9.6)	88.1% (±8.2)	75% (±8.4)
Our proposed method	91% (±4.9)	94.9% (±8.2)	88.2% (±11.3)

more than 13% higher specificity, which means it can predict MDD individuals precisely, and more than 11% accuracy.

One of the most common limitations of machine learning models for classifying MDD individuals from healthy ones is the lack of generalization. The number of records in clinical datasets is usually not favorable for developing a generalizable artificial intelligence model for this task. Because different approaches were employed to collect data (e.g., number of channels, protocols, etc.) in different works, it is difficult to combine various datasets for developing the model. Another limitation of our approach, similar to most deep learning models, is the limited explainability and interpretability, which could be investigated in future works.

VI. CONCLUSION

EEG analysis for the detection of MDD is an arduous and intricate task. Many intelligent models have recruited different methods such as feature extraction, image conservation, and subsampling to achieve high performance. Nonetheless, this study showed that MDD individuals could be classified using EEG time series. Having employed a customized InceptionTime model revealed the power of this deep learning architecture in the face of time series data. In addition, we can accomplish the classification task using a lower number of electrodes. Also, recording first-minute EEG data in the eyes-closed resting-state seems enough to detect MDD. The proposed method can aid healthcare professionals in identifying MDD individuals using fewer EEG channels and shorter recordings.

REFERENCES

[1] C. Arango, C. M. Díaz-Caneja, P. D. McGorry, J. Rapoport, I. E. Sommer, J. A. Vorstman, D. McDaid, O. Marín, E. Serrano-Drozdowskyj, and R. Freedman, "Preventive strategies for mental health," *Lancet Psychiatry*, vol. 5, no. 7, pp. 591–604, 2018.

[2] S. Horan. *Mental Health Summit*. Accessed: Nov. 9, 2018. [Online]. Available: https://www.zevohealth.com/blog/mental-health-summit/

[3] M. Kennis, L. Gerritsen, M. van Dalen, A. Williams, P. Cuijpers, and C. Bockting, "Prospective biomarkers of major depressive disorder: A systematic review and meta-analysis," *Mol. Psychiatry*, vol. 25, no. 2, pp. 321–338, Feb. 2020.

[4] C. Otte, S. M. Gold, B. W. Penninx, C. M. Pariante, A. Etkin, M. Fava, D. C. Mohr, and A. F. Schatzberg, "Major depressive disorder," *Nature Rev. Disease Primers*, vol. 2, no. 1, pp. 1–20, 2016.

[5] D. M. Maurer, T. J. Raymond, and B. N. Davis, "Depression: Screening and diagnosis," *Amer. Family Physician*, vol. 98, no. 8, pp. 508–515, 2018.

[6] F. Edition, "Diagnostic and statistical manual of mental disorders," *Amer. Psychiatric Assoc.*, vol. 21, no. 21, pp. 591–643, 2013.

[7] *Depression and Other Common Mental Disorders: Global Health Estimates*, World Health Org., Brasília, Brazil, 2017.

[8] K. Smith, "Mental health: A world of depression," *Nature*, vol. 515, no. 7526, p. 180, Nov. 2014.

[9] M. Ernst, L. Kallenbach-Kaminski, J. Kaufhold, A. Negele, U. Bahrke, M. Hautzinger, M. E. Beutel, and M. Leuzinger-Bohleber, "Suicide attempts in chronically depressed individuals: What are the risk factors?" *Psychiatry Res.*, vol. 287, May 2020, Art. no. 112481.

[10] R. Lakhan, A. Agrawal, and M. Sharma, "Prevalence of depression, anxiety, and stress during COVID-19 pandemic," *J. Neurosci. Rural Pract.*, vol. 11, no. 4, pp. 519–525, Oct. 2020.

[11] Y. Liu, H. Chen, N. Zhang, X. Wang, Q. Fan, Y. Zhang, L. Huang, B. Hu, and M. Li, "Anxiety and depression symptoms of medical staff under COVID-19 epidemic in China," *J. Affect. Disorders*, vol. 278, pp. 144–148, Jan. 2021.

[12] M. A. Islam, S. D. Barna, H. Raihan, M. N. A. Khan, and M. T. Hossain, "Depression and anxiety among university students during the COVID-19 pandemic in bangladesh: A web-based cross-sectional survey," *PLoS ONE*, vol. 15, no. 8, Aug. 2020, Art. no. e0238162.

[13] K. Allsopp, J. Read, R. Corcoran, and P. Kinderman, "Heterogeneity in psychiatric diagnostic classification," *Psychiatry Res.*, vol. 279, pp. 15–22, Sep. 2019.

[14] M. E. Thase, "Recommendations for screening for depression in adults," *Jama*, vol. 315, no. 4, pp. 349–350, 2016.

[15] T. A. Kato, R. Hashimoto, K. Hayakawa, H. Kubo, M. Watabe, A. R. Teo, and S. Kanba, "Multidimensional anatomy of 'modern type depression' in Japan: A proposal for a different diagnostic approach to depression beyond the DSM-5: Anatomy of Japan's modern depression," *Psychiatry Clin. Neurosci.*, vol. 70, no. 1, pp. 7–23, Jan. 2016.

[16] A. L. Siu, K. Bibbins-Domingo, D. C. Grossman, L. C. Baumann, K. W. Davidson, M. Ebell, F. A. García, M. Gillman, J. Herzstein, and A. R. Kemper, "Screening for depression in adults: U.S. Preventive services task force recommendation statement," *Jama*, vol. 315, no. 4, pp. 380–387, 2016.

[17] A. L. Siu, "Screening for depression in children and adolescents: U.S. Preventive services task force recommendation statement," *Pediatrics*, vol. 137, no. 3, pp. 360–366, Mar. 2016.

[18] J.-J. Chen, S.-J. Bai, W.-W. Li, C.-J. Zhou, P. Zheng, L. Fang, H.-Y. Wang, Y.-Y. Liu, and P. Xie, "Urinary biomarker panel for diagnosing patients with depression and anxiety disorders," *Translational Psychiatry*, vol. 8, no. 1, pp. 1–10, Dec. 2018.

[19] L. Lin, X.-M. Chen, and R.-H. Liu, "Novel urinary metabolite signature for diagnosing postpartum depression," *Neuropsychiatric Disease Treat.*, vol. 13, p. 1263, Jan. 2017.

[20] M. X. Cohen, "Where does EEG come from and what does it mean?" *Trends Neurosci.*, vol. 40, no. 4, pp. 208–218, Apr. 2017.

[21] J. S. Kumar and P. Bhuvaneswari, "Analysis of electroencephalography (EEG) signals and its categorization—A study," *Proc. Eng.*, vol. 38, pp. 2525–2536, Jan. 2012.

[22] P. Olejniczak, "Neurophysiologic basis of EEG," *J. Clin. Neurophysiol.*, vol. 23, no. 3, pp. 186–189, Jun. 2006.

[23] S. Siuly, Y. Li, and Y. Zhang, "Significance of EEG signals in medical and health research," in *EEG Signal Analysis and Classification*. Cham, Switzerland: Springer, 2016, pp. 23–41.

[24] U. R. Acharya, S. V. Sree, P. C. A. Ang, R. Yanti, and J. S. Suri, "Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals," *Int. J. Neural Syst.*, vol. 22, no. 2, Apr. 2012, Art. no. 1250002.

[25] J. F. Cavanagh, A. Napolitano, C. Wu, and A. Mueen, "The patient repository for EEG data+ computational tools (PRED+CT)," *Frontiers Neuroinform.*, vol. 11, p. 67, Nov. 2017.

[26] H. Cai, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao, and Z. Liu, "MODMA dataset: A multi-modal open dataset for mental-disorder analysis," 2020, *arXiv:2002.09283*.

[27] C.-T. Wu, H.-C. Huang, S. Huang, I.-M. Chen, S.-C. Liao, C.-K. Chen, C. Lin, S.-H. Lee, M.-H. Chen, C.-F. Tsai, C.-H. Weng, L.-W. Ko, T.-P. Jung, and Y.-H. Liu, "Resting-state EEG signal for major depressive disorder detection: A systematic validation on a large and diverse dataset," *Biosensors*, vol. 11, no. 12, p. 499, Dec. 2021.

[28] United Nations, *World Population Prospects 2019 Vol (ST/ESA/SE. A/424)*, Dept. Econ. Social Affairs, Population Division, New York, NY, USA, 2019.
[29] W. Mumtaz, L. Xia, M. A. Mohd Yasin, S. S. A. Ali, and A. S. Malik, "A wavelet-based technique to predict treatment outcome for major depressive disorder," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171409.
[30] S. Mahato and S. Paul, "Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry," *J. Med. Syst.*, vol. 44, no. 1, pp. 1–8, Jan. 2020.
[31] E. Aydemir, T. Tuncer, S. Dogan, R. Gururajan, and U. R. Acharya, "Automated major depressive disorder detection using melamine pattern with EEG signals," *Appl. Intell.*, vol. 51, no. 9, pp. 1–18, 2021.
[32] W. Dang, Z. Gao, X. Sun, R. Li, Q. Cai, and C. Grebogi, "Multilayer brain network combined with deep convolutional neural network for detecting major depressive disorder," *Nonlinear Dyn.*, vol. 102, no. 2, pp. 667–677, Oct. 2020.
[33] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach," *Cognit. Neurodynamics*, vol. 15, no. 2, pp. 239–252, Apr. 2021.
[34] H. W. Loh, C. P. Ooi, E. Aydemir, T. Tuncer, S. Dogan, and U. R. Acharya, "Decision support system for major depression detection using spectrogram and convolution neural network with EEG signals," *Expert Syst.*, vol. 39, no. 3, Mar. 2022, Art. no. e12773.
[35] D. M. Khan, N. Yahya, N. Kamel, and I. Faye, "Automated diagnosis of major depressive disorder using brain effective connectivity and 3D convolutional neural network," *IEEE Access*, vol. 9, pp. 8835–8846, 2021.
[36] D. M. Khan, K. Masroor, M. F. M. Jailani, N. Yahya, M. Z. Yusoff, and S. M. Khan, "Development of wavelet coherence EEG as a biomarker for diagnosis of major depressive disorder," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4315–4325, Mar. 2022.
[37] A. Safayari and H. Bolhasani, "Depression diagnosis by deep learning using EEG signals: A systematic review," *Med. Novel Technol. Devices*, vol. 12, Dec. 2021, Art. no. 100102.
[38] T. R. Mullen, C. A. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov. 2015.
[39] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[40] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Ḿuller, and F. Petitjean, "Inceptiontime: Finding alexnet for time series classification," *Data Mining Knowl. Discovery*, vol. 34, pp. 1936–1962, Sep. 2019.
[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.
[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2015, pp. 1-9.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[44] M. Qi, W. Chen, Y. Wang, Z.-M. Ma, and T.-Y. Liu, "Convergence analysis of distributed stochastic gradient descent with shuffling," *Neurocomputing*, vol. 337, pp. 46–57, Apr. 2019.
[45] Y. Nesterov, *Lectures on Convex Optimization*. Cham, Switzerland: Springer, 2018.
[46] A. J. Ferreira and M. A. T. Figueiredo, "Efficient feature selection filters for high-dimensional data," *Pattern Recognit. Lett.*, vol. 33, no. 13, pp. 1794–1804, Oct. 2012.
[47] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.
[48] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika J. Sci. Technol.*, vol. 26, no. 1, pp. 1–12, 2018.
[49] M. Matheny, S. T. Israni, M. Ahmed, and D. Whicher, *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Washington, DC, USA: National Academy of Medicine, 2019.
[50] C. G. Davey, M. Breakspear, J. Pujol, and B. J. Harrison, "A brain model of disturbed self-appraisal in depression," *Amer. J. Psychiatry*, vol. 174, no. 9, pp. 895–903, Sep. 2017.
[51] J. L. Pawluski, J. S. Lonstein, and A. S. Fleming, "The neurobiology of postpartum anxiety and depression," *Trends Neurosci.*, vol. 40, no. 2, pp. 106–120, Feb. 2017.

ALIREZA RAFIEI received the M.Sc. degree in mechatronics engineering from the University of Tehran, Iran, in 2021. He is currently pursuing the Ph.D. degree. His research interests include machine learning applications in medicine and healthcare, time series analysis, digital health, and signal processing.

RASOUL ZAHEDIFAR received the B.Sc. degree in mechanical engineering from Kharazmi University, Tehran, Iran, in 2019. Since then, he has been working as a Researcher in academia with different research groups. His research interests include signal processing, machine learning, mechatronics, and nonlinear control of robots.

CHIRANJIBI SITAULA received the Ph.D. degree from Deakin University, VIC, Australia, in 2021. He has worked in industry and academia in Nepal for a number of years before joining Deakin University for his Ph.D. degree. He is currently a Research Fellow with the Department of Electrical and Computer Systems Engineering, Monash University, VIC. He has published research articles in top-tier conferences and journals in the field of deep learning and machine learning. His research interests include computer vision, signal processing, and machine learning.

FAEZEH MARZBANRAD (Senior Member, IEEE) received the Ph.D. degree from The University of Melbourne, Australia, in 2016. She is currently a Lecturer and the Head of the Biomedical Signal Processing Laboratory, Department of Electrical and Computer Systems Engineering, Monash University, Australia. Her research interests include biomedical signal processing, machine learning, affordable medical technologies, and mobile-health.