DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG

Ayan Seal, *Senior Member, IEEE*, Rishabh Bajpai, Jagriti Agnihotri, Anis Yazidi, *Senior Member, IEEE*, Enrique Herrera-Viedma, *Fellow, IEEE*, and Ondrej Krejcar

***Abstract*— Depression is a common reason for an increase in suicide cases worldwide. Thus, to mitigate the effects of depression, accurate diagnosis and treatment are needed. An electroencephalogram (EEG) is an instrument used to measure and record the brain's electrical activities. It can be utilized to produce the exact report on the level of depression. Previous studies proved the feasibility of the usage of EEG data and deep learning (DL) models for diagnosing mental illness. Therefore, this study proposes a DL-based convolutional neural network (CNN) called DeprNet for classifying the EEG data of depressed and normal subjects. Here, the Patient Health Questionnaire 9 score is used for quantifying the level of depression. The performance of DeprNet in two experiments, namely, the recordwise split and the subjectwise split, is presented in this study. The results attained by DeprNet have an accuracy of 0.9937, and the area under the receiver operating characteristic curve (AUC) of 0.999 is achieved when recordwise split data are considered. On the other hand, an accuracy of 0.914 and the AUC of 0.956 are obtained, while subjectwise split data are employed. These results suggest that CNN trained on recordwise split data gets overtrained on EEG data with a small number of subjects. The performance of DeprNet is remarkable compared with the other eight baseline models. Furthermore, on visualizing the last CNN layer, it is found that the values of right electrodes are prominent for depressed subjects, whereas, for normal subjects, the values of left electrodes are prominent.**

***Index Terms*— Convolutional neural network (CNN), electroencephalography, measurement of depression, pattern classification, visualization.**

Manuscript received November 2, 2020; revised December 23, 2020; accepted January 11, 2021. Date of publication January 25, 2021; date of current version February 8, 2021. This work is partially supported by the project "Prediction of diseases through computer assisted diagnosis system using images captured by minimally-invasive and non-invasive modalities", Computer Science and Engineering, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur India (under ID: SPARC-MHRD-231). This work is also partially supported by the project "Smart Solutions in Ubiquitous Computing Environments", Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic (under ID: UHK-FIM-GE-2021); project at Universiti Teknologi Malaysia (UTM) under Research University Grant Vot-20H04, Malaysia Research University Network (MRUN) Vot 4L876 and the Fundamental Research Grant Scheme (FRGS) Vot5F073 supported under Ministry of Education Malaysia for the completion of the research. The Associate Editor coordinating the review process was Adam G. Polak. *(Corresponding author: Ayan Seal.)*

Ayan Seal is with the PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India, and also with the Center for Basic and Applied Science, Faculty of Informatics and Management, University of Hradec Králové, 500 03 Hradec Králové, Czech Republic (e-mail: ayanseal30@ieee.org).

Rishabh Bajpai and Jagriti Agnihotri are with the PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India (e-mail: rishabhbajpai24@gmail.com; jagriti.agni@gmail.com).

Anis Yazidi is with the Research Group in Applied Artificial Intelligence, Oslo Metropolitan University, 460167 Oslo, Norway (e-mail: anisy@oslomet.no).

Enrique Herrera-Viedma is with the Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain, and also with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: viedma@decsai.ugr.es).

Ondrej Krejcar is with the Center for Basic and Applied Science, Faculty of Informatics and Management, University of Hradec Králové, 500 03 Hradec Králové, Czech Republic, and also with the Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia (e-mail: ondrej.krejcar@uhk.cz).

Digital Object Identifier 10.1109/TIM.2021.3053999

I. Introduction

MENTAL illness, also known as mental health disorders, is a physical illness of the brain that might affect the thinking process, behavior, and mood. It also leads to loss of interest and energy, may cause adverse effects on relationships, performance at the workplace, and increase the risk of suicide. Worldwide, almost 13% of the child population, 46% of adolescents, and 19% of the adult population struggle with mental illness each year [1]. Thus, the diagnosis of depression in the early curable stages is crucial to prevent it from reaching a severe and irreversible state and to save the life of depressed individuals. Generally, the symptoms of depression are reflected in the behavior of the patient. Hence, doctors conduct talking sessions and use questionnaires as screening tools for determining the level of depression. However, the outcome of a talking session is dependent on the psychiatrist's or counselor's proficiency. Moreover, depressed patients are less likely to seek help due to the stigma attached to mental illness. As a result, a significant number of depressed individuals do not get the best possible medication and sufficient revival time. Thus, finding suitable and efficient approaches for detecting depression is an emerging field of study, and the recent developments in the instrument or sensor technology open up new horizons to diagnose depression. Among electroencephalogram (EEG), magnetoencephalography, magnetic resonance imaging, functional magnetic resonance imaging, and physiological data, EEG is a portable technology that can capture the electrical activity of brain neurons from the scalp surface in real time. It is observed that most of the cognitive behavior and psychological activities are analyzed by EEG [2] because the EEG signal acquired from the parietal lobe of the human brain is related to the cognitive tasks and emotional states [3]. Thus, the EEG signal could be

exploited in order to understand the human cognitive process and diagnose mental illness. However, it is difficult to interpret nonstationary, nonlinear, and complex EEG signal visually. Moreover, it is a tedious task to fetch task-relevant features from the EEG signal. Naturally, linear methods cannot observe the complex dynamic variations in the EEG signal. Therefore, the deep learning (DL)-based approaches could be used to extract features from the EEG signal for computer-aided diagnosis (CAD) of depression because DL-based methods can extract extremely complex and highly nonlinear features automatically from raw data with little or no effort [4]–[6].

Neuroscience, psychology, and cognitive science researchers have analyzed EEG data extensively in various aspects. However, Craik *et al.* [4] revealed that 37% of the previous studies did not preprocess the EEG data, 49% removed artifacts manually, and 14% used automatic artifact removal techniques. The study also presented that 41% of previous studies considered calculated features, 39% employed signal values, and 20% used images, which were transformed from EEG data as input data for the network. In addition, for network architecture, 53% exploited convolutional neural network (CNN)-based models, 18% explored the deep belief network (DBN), 10% used the recurrent neural network (NN), 11% employed multilayer perceptron-based models, and rest 8% considered the stacked autoencoder. The study suggested that, due to less preprocessing, CNN is the favorite choice of the scholars dealing with EEG data.

A. Motivation and Contributions

It is clear from the literature that, primarily, two feature extraction techniques, namely, manual and automatic, were explored for identifying depression using EEG. However, the accuracies reported by most of the previous studies are not satisfactory. Thus, these methods are not suitable for practical use. The previous studies using DL models completely missed the spatial information for the classification as single-channel raw EEG data were considered as input for the network [7]. However, the performance of CNN-based methods can be improved by selecting the hyperparameters of the architecture [8]. It motivates us to work further in this direction. This study aims to design a simple CNN, which considers both spatial information and temporal information for attaining high classification accuracy. The simple network architecture helps in understanding the role of the left and right hemispheres activities of the brain for the classification of depression. The salient contributions of the proposed study are listed as follows.

1) A novel CNN framework called DeprNet is introduced for classifying depressed and normal subjects based on short EEG recordings of 4 s from 19 channels. The use of short EEG recording facilitates the deployment of the proposed method in practical scenarios.
2) This study includes 17 307 samples and obtains 0.914 accuracy, which is the highest among the other similar CNN-based architectures. In addition, other quantitative classification metrics, namely, precision, recall, and F1-score, are considered while comparing the performance of the proposed method with the results of some of the state-of-the-art methods due to the accuracy paradox.
3) Furthermore, this study presents that the consequences of depression are different in the activities of the right and the left hemispheres of the brain. This conclusion is drawn by analyzing and visualizing the last layer of the proposed CNN and showing that this layer uses the values of right electrodes for depressed subjects and left electrodes for the nondepressed subjects while detecting the level of depression.

The structure of this article is given as follows. A brief review of some of the existing works related to depression classification is discussed in Section II. The proposed method is illustrated in Section III. Section III includes the obtained results of the DeprNet along with other baseline methods. The visualization of the deepest DeprNet's layer is also presented in the same section. Finally, the conclusion of our work is drawn in Section V.

II. Related Work

All the related works can broadly be divided into two categories: handcrafted feature-based methods and raw data-based methods. This section presents the works related to both the categories, sequentially. The first notable contribution in this field came from Puthankattil and Joseph [9] in 2012. They considered 15 depressed and 15 nondepressed subjects for the training of a two-layer feedforward artificial NN. They used all frequency ranges of EEG signals for the extraction of relative wavelet energy and signal entropy-based features. Since this was a preliminary study, a small set of features and a conventional machine learning technique were employed for classification. In the same year, Ahmandlou *et al.* [10] exploited nonlinear features, wavelet filter banks, and fractal dimensions with enhanced probabilistic NN. Their results of Higuchi's fractal dimensions showed higher complexity of left, right, and overall frontal lobes of the brain of people with major depressive disorder (MDD) compared to non-MDD in beta and gamma subbands. Hosseinifard *et al.* [11] also confirmed that nonlinear features are significantly effective for analyzing EEG data. They used a large data set of EEG recordings of 90 subjects (45 normal and 45 depressed) and compared three machine learning algorithms, namely, logistic regression, linear discriminant analysis, and k-nearest neighbor (KNN) for classification. They displayed that, among nonlinear features, the correlation dimension was a powerful feature for analyzing EEG signals while identifying depressed and nondepressed subjects. However, a combined model of linear and nonlinear features can provide better recognition accuracies. Faust *et al.* [12] exploited wavelet packet decomposition and other nonlinear features with PNN and compared the left electrodes results with right electrodes results. They also compared the results of seven classical classification algorithms. Despite the great accuracy of the models, there are some flaws in the methods. This study had not considered redundancy of the features and ignored the step of feature selection. Also, the reported high accuracy might be the result of overfitting

as the data were divided into training and testing groups in a recordwise manner. Moreover, Acharya *et al.* [13] employed nonlinear methods with a support vector machine (SVM) for classification and developed a depression diagnosis index. They also introduced a depression diagnosis index through a judicious combination of the nonlinear features. Since there is no evidence of the relation of nonlinear features used in the study for defining depression diagnosis index with depression, the use of depression diagnosis index for the classification is questionable. Similarly, an investigation was conducted by Cai *et al.* [14] on 178 subjects using three-electrode pervasive EEG collectors. This study showed that the DBN performs better than the traditional shallow models [14]. Although the accuracy of the model was 78.24%, it showed the feasibility of the usage of a small pervasive EEG collector for the detection of depression. However, in order to understand the generalizability of the method, its performance should be tested on other electrodes as well. Mumtaz *et al.* [15] employed alpha interhemispheric asymmetry and power of frequency bands with SVM. Many other researchers also tested the importance of the alpha asymmetry and the power of different frequency bands in detecting depression. Liao *et al.* [16] used SVM and performed principal component analysis to extract features. Bairy *et al.* [17] adopted linear prediction coding with the bagged tree. Sharma *et al.* [18] examined the role of handcrafted features and SVM. Cai *et al.* [19] considered 213 subjects' data for comparing the classification performance of four algorithms and achieved the best results with the KNN classifier. Although the number of subjects used in this study is significantly higher than other studies, the maximum average accuracy attained by the method is 76.98%. Their recent work [20] compared the performance of KNN, DT, and SVM on the same data set. Their KNN model obtained the highest accuracy of 89.98%. Shen *et al.* [21] presented an improved empirical mode decomposition (EMD) applying singular value decomposition (SVD)-based feature extraction method that can fetch the features coefficients of expansion using all intrinsic mode functions accurately.

Since depression affects both superficial and deeper structures of the brain [22], handcrafted features may not capture most of the depression-related effects. However, DL-based approaches can be better for understanding the brain of depressed people. Acharya *et al.* [23] described a model on screening of depression by EEG signals using DL. They showed that the EEG signals from the left hemisphere of the brain are less distinctive than those from the right hemisphere. Since previous literature suggested that depression affects the whole brain, the performance of the presented networks can be increased by considering more EEG channels for training. Ay *et al.* [24] exploited long short-term memory (LSTM) networks and obtained a relatively higher performance than other methods on the same data set. They considered a 1-D input vector containing EEG recording of a single channel for classification. However, due to the possibility of overfitting of their model, it cannot be considered for practical application. As they trained and tested the model on data of the same subjects, their model will not perform satisfactorily on untrained subjects. This limits the usability of their method.

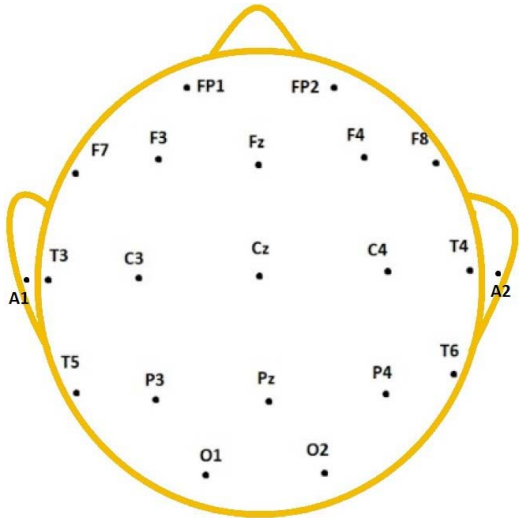

Fig. 1. Positions of 19 electrodes used in data set according to the 10–20-electrode positioning system [28], [29].

Recently, Li *et al.* [25] employed a spectrogram as input to the network and trained a CNN model. They targeted mild depression recognition in clinical practice; thus, they introduced a CAD system using CNN. Their method considered 128 channels' EEG recording for classification. Furthermore, this study used both special and temporal information of EEG data for classification. However, placement time for 128 electrodes is huge and not all clinics can afford such a costly system. There is a scope for improving the performance of the network. Shah *et al.* [26] exploited spiking NN-based architecture named NewCube. They used resting-state EEG collected from 22 subjects, including healthy and mild-depressed patients. They further exploited the multilayer perceptron for comparative analysis. Thoduparambil *et al.* [27] designed a model comprised of CNN and LSTM layers to learn local characteristics and the EEG signal sequence, respectively.

III. Methodology

The methodology used for quantifying and comparing the performance of DeprNet with other baseline methods is explained in this section.

A. Data Set Description

In this study, we use resting-state EEG recording of 33 subjects with a duration of 9 minutes each because the effects of depression are reflected in the resting state. Out of 33 subjects, 18 subjects are normal, and 15 subjects are depressed. Moreover, the Patient Health Questionnaire 9 (PHQ-9) score of each subject is computed by conducting an interview session and consulting a psychologist. According to 10–20-electrode positioning system (as shown in Fig. 1) [28], [29], 19 channels with mean of ear electrodes (A1 and A2) as common references are considered while developing the data set.

During the recording, a 0.1-Hz high-pass filter, a 100-Hz low-pass filter, and a 50-Hz notch filter are considered to

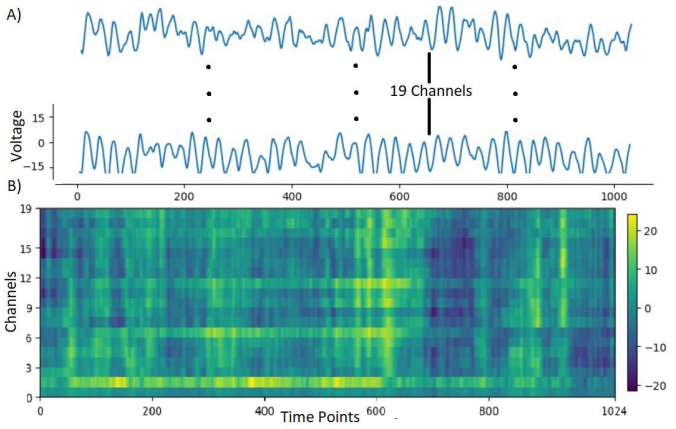

Fig. 2. Pictorial representation of raw EEG data used as input by the proposed method. (a) Sample time-series plot of a data sample showing raw EEG data of 19 channels and duration of 4 s. The y-axis corresponds to voltage values of the signal (μV), and the x-axis corresponds to time points (4 s = 1024 time points). (b) Sample input matrix is shown, which is constructed from the values of all channels and time points of one sample. The y-axis corresponds to channels, the x-axis corresponds to time points, and the color corresponds to the voltage values (μV). The normalized version of this image is then used as an input image of the network.

remove the low-frequency noise, irrelevant signals, and the baseline noise from the data, respectively. The independent component analysis (ICA) is performed on the signals of open eye state to extract and remove artifacts related to eye movements [30]. The ICA assumes that the signal can be thought of as a weighted sum of statistically independent non-Gaussian components [31]. The ICA is applied to the signal using the EEGLAB toolbox with the "runica" algorithm with its default settings. After preprocessing the data of all subjects, a total of 17 310-s EEG recording is considered for further analysis. The data is split into 17 307 data points, and two data matrices of different dimensions 17 307 × 19 × 1.024 and 17 307 × 1 are constructed. The first matrix is the input data matrix, whereas the second one is the output data matrix (contained class of subjects). In the input data matrix, each data point has data of 4 s (with 75% overlapping) and 19 channels. As the sampling frequency of the data is 256 Hz, the 4-s recording has 1024 (4 × 256) time points (see Fig. 2). In the output data matrix, each data point has a value of either "0" or "1." "0" indicates that the data point belongs to a nondepressed subject, while "1" is used to represent the data point that belongs to a depressed subject. The categorization of each subject is made based on the PHQ-9 score and the expertise of a psychologist. The reasons for choosing a time window of 4 s with 75% overlapping are as follows.

1) Four seconds have 1024 samples, which are sufficient for representing frequency components up to 512 Hz (Nyquist frequency). Thus, the network can extract all the essential features.
2) The data set of 33 subjects is divided into 17 307 data points, which are enough for training, testing, and validating a binary classifier.
3) The 75% overlapping can remove any phase shift effect in the data and also increases the number of data points without using any other data augmentation technique.

B. Architecture of DeprNet

The proposed CNN, DeprNet, consists of five convolutional layers, five batch normalization layers, five max-pooling layers, and three fully connected layers. The last fully connected layer uses softmax activation function, while all other layers use leaky rectified linear unit (LeakyReLU) activation function [32], [33]. The arrangement of these layers is shown in Fig. 3. The details of parameters and filters are reported in Table I. Even though the input data are 2-D, the network performs a 1-D convolution operation. The 1-D convolution is applied on the time dimension, i.e., y-axis, and it ensures that the information associated with the spatial dimension, i.e., x-axis, remains as it is. This method of keeping the spatial information separated is useful in understanding how spatial information is processed by the network and which channels are important for the detection of depression (explained in Section III-C).

1) Convolutional Layers: Convolution is an operation applied on a signal to get transformed and relevant signals. For convolution, a filter or linear and time-invariant is required, which is multiplied and translated on the signal to get new signal values. In DeprNet, the first convolutional layer is preceded by the input layer. The number of filters used is either 128 or 64 or 32 for all layers. The first three convolutional layers are convolved with a filter size of 1×5. The filter sizes in the fourth and fifth convolutional layers are 1×3 and 1×2, respectively. When convolution is implemented using a NN, a kernel/filter is slid over the input signals to obtain an output, which is also known as the activation map of the layer [34]. We choose 128 filters for C1 to extract most of the essential low-level features and reduce the number of features with the depth of the network to convert the EEG data into a low-dimensional space. The C5 layer only contains important high-level relevant features.

2) Batch Normalization: The batch normalization is adopted to stabilize the network by normalizing the output of the previous layer [35]. In DeprNet, five batch normalization layers are included, one layer after each convolutional layer. The network uses the LeakyReLU activation function on the output of the batch normalization layers. We find a better convergence rate when the batch normalization is employed than when batch normalization is not considered.

3) Pooling Layers: In DeprNet, max-pooling layers are exploited for down-sampling the data [36]. The network applies 1-D pooling operations on the time dimension. The filter size of all five layers is equal to 1×2.

4) Fully Connected Layers: After five rounds of convolution, batch normalization, and pooling layers, two dense layers are kept. The first layer has 16, while the second consists of eight neurons. After visualizing the activation map of the 15th layer, we find that CNN layers can extract essential features for detecting the depression. Therefore, we use a fewer number of neurons in the fully connected layers (as shown in Section IV-G).

5) Softmax Layer: The last layer that predicts the final output is a dense layer with a softmax activation function. The softmax function gives a vector as output that represents the probability distributions of a list of potential

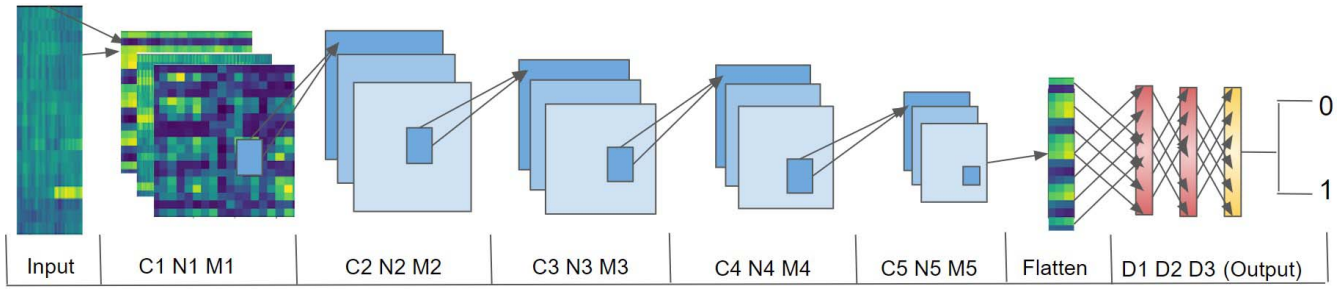

Fig. 3. Architecture of the proposed DL-based CNN model, DeprNet.

TABLE I
ARCHITECTURE OF DEPRNET

S.No	Layer Name	Layer Type	Filter size	Layer parameters	Stride x	Stride y	Output size	Parameters
1	C1	convolutional 1	1×5	128 filters	1	2	19×1020×128	768
2	N1	Batch_normalization 1					19×1020×128	512
3	M1	Max_pooling 1	1×2		1	2	19×510×128	
4	C2	convolutional 2	1×5	64 filters	1	2	19×506×64	41024
5	N2	Batch_normalization 2					19×506×64	256
6	M2	Max_pooling 2	1×2		1	2	19×253×64	
7	C3	convolutional 3	1×5	64 filters	1	2	19×249×64	20544
8	N3	Batch_normalization 3					19×249×64	256
9	M3	Max_pooling 3	1×2		1	2	19×124×64	
10	C4	convolutional 4	1×3	32 filters	1	2	19×122×32	6176
11	N4	Batch_normalization 4					19×122×32	128
12	M4	Max_pooling 4	1×2		1	2	19×61×32	
13	C5	convolutional 5	1×2	32 filters	1	2	19×60×32	2080
14	N5	Batch_normalization 5					19×60×32	128
15	M5	Max_pooling 5	1×2		1	2	19×30×32	
16	D1	Dense 1		16 neurons			16	291856
17	D2	Dense 2		8 neurons			8	136
18	D3	Dense 3 (softmax)		2 neurons			2	18

outcomes [37], [38]. In this case, there are two potential outcomes; thus, we need two neurons to represent them.

6) Loss Function: The loss function considered in the DeprNet network is binary cross-entropy loss, J(Θ) [38], for binary classifiers, which is computed using the following equation:

$$J(\Theta) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \qquad (1)$$

where $\hat{y}_i$ and y_i are the predicted label and the actual class label of the ith sample, respectively, and N is the total number of samples. The model is trained over multiple iterations to minimize the value of J(Θ).

C. Visualization of Features Learned by Convolutional Layers

The convolutional layers and fully connected layers of a NN are responsible for extracting features and performing classification, respectively. Since the number of neurons in the last three layers is small ($16 + 8 + 2 = 26$), most of the decisions related to classification are dependent on convolutional layers. This suggests that the last layer of CNN may contain useful information about the decisions the network takes while classifying the subjects. Therefore, this study includes the visualization of the activation map of the max-pooling layer, M5 [39]–[41].

For this analysis, the network is trained on all data points, i.e., 17 207 samples. After training the network, the data points are feedforwarded to the network to get their activation maps. For each data point, the output volume of size $19 \times 30 \times 32$ is obtained from layer 15, which is then converted into a vector of size 19 by taking the average of responses across 32 filters and 30 time points. We further average the responses of all data points corresponding to a subject and, finally, construct response vectors having 19 values for each subject. The method explained above gives average values of 19 channels of filtered data (filtered from all convolutional layers) for each subject. These values can suggest which channels are more responsible for depression. For better presentation, these response vectors are plotted as heat maps (see Fig. 4). From these 19 values, eight, eight, and three values correspond to the left-hand side, right-hand side, and middle electrodes, respectively. A parameter denoting skewness of the spatial distribution of values of the electrodes for each subject is computed by the following equation:

$$\alpha = L - R \qquad (2)$$

where L and R are the values of the left and right electrodes of the response vector, respectively. A positive value of α denotes that the values of the response vector of left electrodes are more than of the response vector of right electrodes. Similarly, the negative value of α denotes that the values of the response

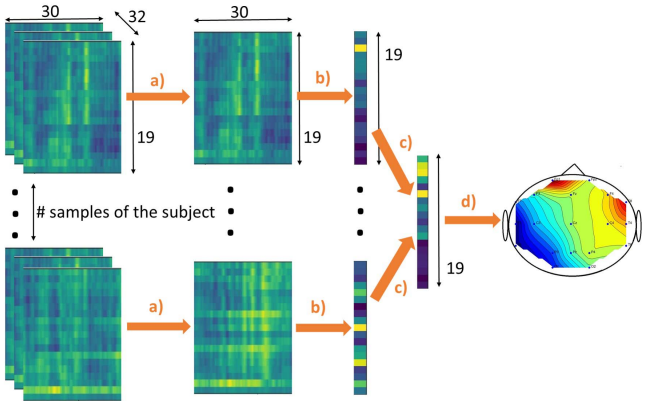

Fig. 4. Process of computing heat map for a subject is explained in this figure. Step a) Averaging the convoluted volume across filters. The convoluted volume extracted from the 15th layer of DeprNet is averaged across its filters. Step b) Averaging across 30 time points. This averaging results in the response vector of a sample. Step c) Averaging across all sample points of subjects. This output vector is referred to as the response vector of the given subject. Step d) Conversion of response vector into a heat map. The values' response vector is normalized and mapped to their corresponding locations in the channel space.

TABLE II
SPECIFICATION OF THE SYSTEM

Name	Parameter
GPU RAM	16GB
System RAM	13GB
CPU	Intel(R) Xeon(R)@ 2.30GHz
Graphics processor	Tesla P100-PCIE
Cuda cores	2560
Memory type	GDDR5X
Language	Python 3.8.1

vector of the left electrodes are less than that of the response vector of the right electrodes.

IV. Experimental Results and Discussion

A. Environment Settings

In this study, the Keras framework and Google Colab platform are used for the training and testing of a DL model. Python 3.8.1 language is employed since many of the DL libraries are developed using it. The specification of the system is noted in Table II.

B. Evaluation Metrics

Accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) are considered for evaluating the performance of the DeprNet. The definition of each of these evaluation metrics is beyond the scope of this study. However, the interested readers can refer to [42]–[44] to know about these parameters in detail. The AUC is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: true positive rate/recall and false-positive rate. The value of each of these metrics lies between 0 and 1. A higher value close to 1 is expected. A higher value of each of these metrics signifies the better performance of a model. The AUC is considered in this study to visualize the performance of a model graphically [45].

C. Computational Protocol

The classification report of the DeprNet is compared with three recent DL-based models, namely, AchLSTM: automated depression detection using deep representation and sequence learning with EEG signals [24]; AchCNN: automated EEG-based screening of depression using deep CNN [23]; and T-LSTM: EEG-based DL model for the automatic detection of clinical depression [27], and five recent handcrafted feature-based methods, namely, H-KNN1: a pervasive approach to EEG-based depression detection [19]; H-KNN2: feature-level fusion approaches based on multimodal EEG data for depression recognition [20]; S-EMD: an improved EMD of electroencephalogram signals for depression detection [21]; S-SVM: automated detection of abnormal EEG signals using localized wavelet filter banks [46]; and H-DBN: pervasive EEG diagnosis of depression using DBN with three-electrode EEG collector [14]. The scholars have not used any names to refer to their works, so we provide meaningful names. The detailed descriptions of these models are given as follows.

1) AchLSTM [24] was based on a hybrid network of CNN and LSTM layers, and it acquired the highest classification accuracy, among the other similar methods. That makes it an obvious choice for comparison. The architecture of AchLSTM consisted of eleven layers. The first, second, fifth, and sixth layers were convolutional layers with a filter size of 5, 3, 13, and 7, respectively, with the number of filters as 64, 128, 128, and 32, respectively. The third layer was a max-pooling layer with a filter size of 2 and a stride of 2. The fourth and tenth layers were dropout layers with a dropout rate of 0.2. The eighth layer was a flatten layer used for flattening the data volume received from the LSTM layer. The ninth and eleventh layers were dense layers with the number of neurons equal to 64 and 2, respectively. The last layer employed softmax as an activation function and other layers considered ReLU as an activation function. Since the performance reported in the original paper of AchLSTM was better for right side electrodes, we train it on Fp2-T4 electrodes during its implementation. First, the data are converted into data points, where each data point contains EEG data of 2000 sample points. Then, these data points are scaled to a range of [0, 1]. Then, the network is trained with the number of epochs of 15 and a batch size of 32. These values are considered for training after trying many other combinations.
2) AchCNN [23] is the most relevant study for our network, as its architecture is similar to DeprNet. Depression affects various regions of the brain differently. Therefore, a multidimensional analysis, including time-varying activities of various regions of the brain, is needed to capture the effects of depression on the brain. The AchCNN only considered temporal information of EEG signals, while DeprNet uses both spatial

and temporal information for predicting the level of depression. Moreover, the working of DeprNet was more focused on extracting significant information from EEG signals in the form of features, while the working of AchCNN is more focused on using more number of fully connected neurons for the classification. The data points used in this method are the same as the AchLSTM. The AchCNN was made up of five convolutional layers, five pooling layers, and three dense layers. The first layer was a convolutional layer, and each convolutional layer was followed by a max-pooling layer. Each convolutional layer considered a filter of size 5 and stride of 1 and each max-pooling layer used a filter of size 2 and stride 2. Then, the network was trained with a number of epochs of value 25 and a batch size of 64. These values were considered for training after trying many other combinations.

3) T-LSTM [27] also implemented CNN and LSTM units for the detection of depression. T-LSTM had three CNN layers each followed by a max-pooling layer of filter size 2. The three CNN layers had 128×7, 64×5, and 32×3 filter sizes. These CNN layers were followed by two LSTM layers with a unit size of 2 and two dense layers with unit sizes of 10 and 2. These dense layers had a dropout rate of 0.2. First, the raw EEG data are preprocessed by removing the eye blinks from the z-normalized data. The preprocessed data are then passed to T-LSTM as the input layer.
4) Since H-KNN1 [19] used data from 213 subjects for designing the method, it is worthy of considering it for comparison. Fp1, Fp2, and Fpz electrodes were considered by Cai *et al.* [19] in H-KNN. However, the value of Fpz electrodes is interpolated in our analysis because the reading of Fpz electrodes is not available in our data set. First, the data are divided into 6-s epochs; then, 270 handcrafted combining linear and nonlinear features from all frequency range, namely, alpha, beta, gamma, delta, and theta, are extracted. These features are z-normalized before training the model. The interested reader can refer to [19] to see the process of extraction of these features. The value of the number of neighbors for H-KNN is tuned, and the best accuracy is obtained at 50.
5) H-KNN2 [20] exploited 60 linear features and 36 nonlinear features of EEG signals in the whole band (0.5–50 Hz), theta, alpha, beta, and gamma bands. The parameters of KNN used are the number of neighbors as 3, the algorithm as ball tree, and the metric as the Euclidean distance. The interested reader can refer to [20] to know the process of extraction of these features.
6) *S-EMD [21]:* The EEG data of three channels (Fp1, Fp2, and Fpz) are preprocessed, as explained in [21]. Then, an improved EMD applying the SVD-based feature extraction method is used for extracting the features coefficients of expansion based on all intrinsic mode functions. The parameters and methodology of the modified EMD feature extraction technique used the same that is discussed in [21]. Then, the SVM algorithm is exploited for classifying the signals into depressed and healthy classes.
7) S-SVM [46] was exploited for automated detection of epilepsy using stopband energy (SBE) minimized orthogonal wavelet filter bank. S-SVM is tested for depression classification by considering the methods, as mentioned in [46]. The EEG data are split into channels, followed by the preprocessing of the data of T5-O1 and F4-C4. Then, the two-band SBE minimized orthogonal wavelet filter bank for the wavelet decomposition is exploited for decomposing the EEG signal [46]. Then, fuzzy entropy, fractal dimension features, and logarithmic squared are extracted from the data (as explained in [46]). Then, the SVM model is exploited for classification purposes.
8) H-DBN [14] is considered for comparison due to its robustness reported by the researchers. The data are divided into 6-s epochs, then 85 linear and nonlinear features are extracted from theta, alpha, beta, and gamma frequency range, and then, the data are normalized. We consider other parameters from the original paper. The interested reader can refer to [14] to see the process of extraction of these features.

D. Training and Testing of the Network

Two experiments are conducted based on recordwise and subjectwise data splitting.

1) *Recordwise Split Data:* For training, validation, and testing, the data are randomly divided into three parts. Training has 53.3%; validation has 13.4%; and testing has 33.3% of the whole data. In other words, the network is trained on 9230 samples, validated on 2308 samples, and tested on 5769 samples.
2) *Subjectwise Split Data:* For training, validation, and testing, the data of 33 subjects are divided into ten groups such that no two groups have a common subject. Among these groups, seven groups contain data of three subjects each, and the remaining three groups contain data of four subjects each. During the division, it is ensured that each group contains at least one subject from each category of depression, i.e., no depression, mild depression, and severe depression. By doing this, it is tried to divide subjects into groups uniformly. Then, the network is trained with tenfold cross-validation. One group is used as the testing set in each fold, and the remaining nine groups are used as training and validation sets. The remaining nine groups' data points are randomly shuffled before splitting into two parts of 80% (training set) and 20% (validation set).

For all nine methods mentioned in Section IV-C, including DeprNet, tenfold cross-validation with the same data set was implemented. Even for each fold, the same training, validation, and testing sets are used across all methods including DeprNet. In other words, the data set of 33 subjects is divided into ten groups in a subjectwise manner, and all methods in tenfold cross-validation use these same ten groups.

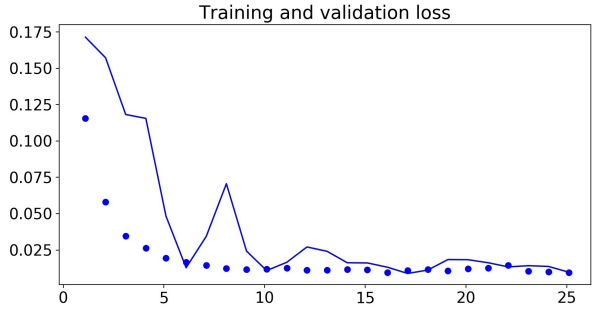

Fig. 5. Training of DeprNet on recordwise split data: training (dotted) and validation (line) loss (y-axis) with training epochs (x-axis).

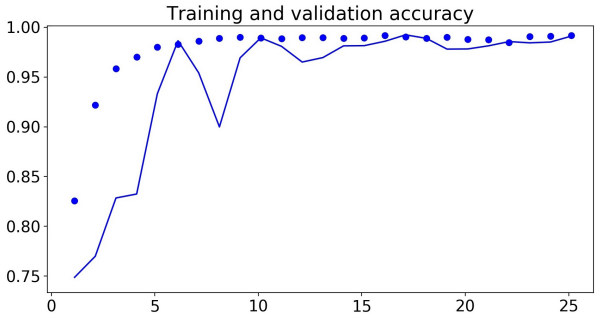

Fig. 6. Training of DeprNet on recordwise split data: training (dotted) and validation (line) accuracy (y-axis) with training epochs (x-axis).

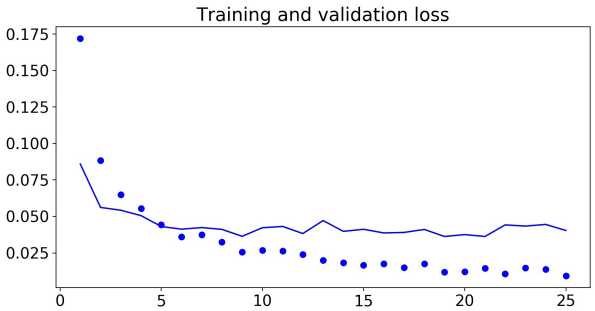

Fig. 7. Training of DeprNet on subjectwise split data: training (dotted) and validation (line) loss (y-axis) with training epochs (x-axis).

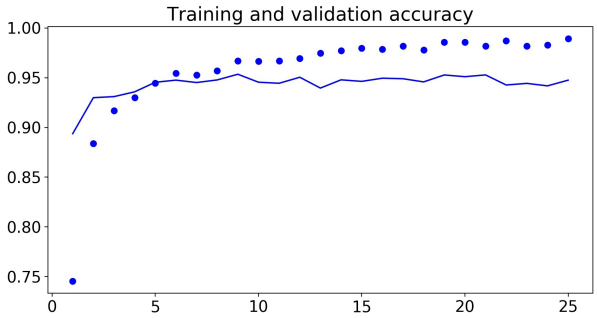

Fig. 8. Training of DeprNet on subjectwise split data: training (dotted) and validation (line) accuracy (y-axis) with training epochs (x-axis).

The Adam optimizer, with a learning rate of 0.0005, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$, is used for updating the weight of the network [47]. β_1 and β_2 are the exponential decay rates of the first- and second-moment estimates of the Adam optimizer algorithm, respectively [47]. Before the actual training of the model, we manually verify that the model typically converges in about 20 epochs. Thus, we use the value of a maximum number of training epochs as 25. Equation 1 is considered for calculating training loss and validation loss during the training. A model checkpoint of early stopping with seven epoch points on validation loss is employed to avoid overfitting of the model. The network is trained with a batch size of 64 and with 25 as the number of epochs. For recordwise split data, the training and validation loss and accuracy of the network during the training are shown in Figs. 5 and 6, respectively. For subjectwise split data, the training and validation loss and accuracy of the network during the training are averaged across the folds shown in Figs. 7 and 8, respectively.

E. Results and Comparison

The classification reports of the abovementioned eight baseline methods and DeprNet on recordwise split data and subjectwise split data are reported in Tables III and IV, respectively. Figs. 9 and 10 show the ROC and value of AUC of each of the above-said methods, including DeprNet on recordwise split data and subjectwise split data, respectively. For subjectwise split data, data from all folds are considered for computing the values of the ROC curves. These results suggest that CNN trained on recordwise split data gets overtrained on EEG data with a small number of subjects. Furthermore, results also illustrate that DeprNet is outperforming other methods on our data set.

TABLE III

EXPERIMENT RECORDWISE SPLIT DATA: THE CLASSIFICATION REPORT OF ACHCNN, ACHLSTM, T-LSTM, H-KNN1, H-KNN2, S-EMD, S-SVM, H-DBN, AND DEPRNET

	Precision	Recall	F1-score	Accuracy
AchCNN [23]	0.661	0.628	0.644	0.692
AchLSTM [24]	0.677	0.791	0.730	0.740
T-LSTM [27]	0.868	0.835	0.851	0.837
H-KNN1 [19]	0.567	0.636	0.599	0.623
H-KNN2 [20]	0.683	0.731	0.706	0.730
S-EMD [21]	0.784	0.740	0.761	0.742
S-SVM [46]	0.629	0.694	0.660	0.683
H-DBN [14]	0.605	0.687	0.643	0.662
DeprNet	0.994	0.991	0.993	**0.9937**

TABLE IV

EXPERIMENT SUBJECTWISE SPLIT DATA: THE CLASSIFICATION REPORT OF ACHCNN, ACHLSTM, T-LSTM, H-KNN1, H-KNN2, S-EMD, S-SVM, H-DBN, AND DEPRNET

	Precision	Recall	F1-score	Accuracy
AchCNN [23]	0.581	0.639	0.610	0.681
AchLSTM [24]	0.613	0.913	0.734	0.744
T-LSTM [27]	0.780	0.818	0.799	0.817
H-KNN1 [19]	0.527	0.655	0.588	0.641
H-KNN2 [20]	0.766	0.725	0.745	0.724
S-EMD [21]	0.667	0.735	0.699	0.720
S-SVM [46]	0.617	0.680	0.647	0.671
H-DBN [14]	0.529	0.778	0.627	0.642
DeprNet	0.919	0.887	0.895	**0.914**

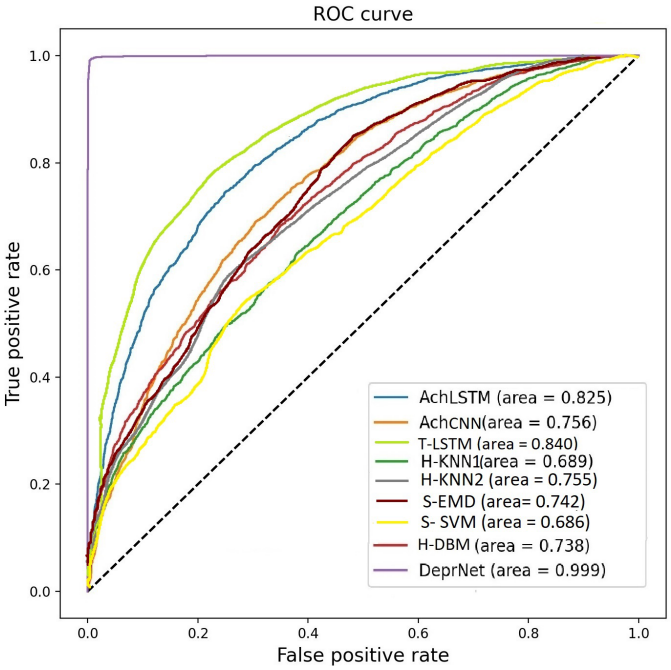

Fig. 9. Experiment recordwise split data: ROC curve of all nine networks: violet: DeprNet; blue: AchLSTM [24]; orange: AchCNN [23]; light Green: T-LSTM [27]; green: H-KNN1 [19]; purple: H-KNN2 [20]; brown: S-EMD [21]; yellow: S-SVM [46]; and red: H-DBM [14].

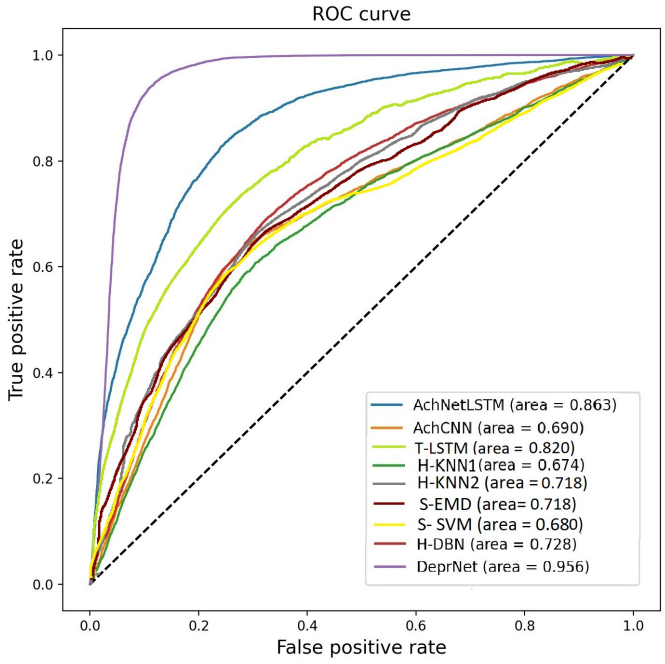

Fig. 10. Experiment subjectwise split data: ROC curve of all nine networks: violet: DeprNet; blue: AchLSTM [24]; orange: AchCNN [23]; light green: T-LSTM [27]; green: H-KNN1 [19]; purple: H-KNN2 [20]; brown: S-EMD [21]; yellow: S-SVM [46]; and red: H-DBM [14].

There are three possible reasons for the better performance of DeprNet than other networks. First, other methods failed to use both spatial and temporal dimensions of the EEG data, while DeprNet used both dimensions during the training. Other methods either considered data of single-channel or employed only a few channels for classification. In other words, they did not consider spatial information of EEG data appropriately. Moreover, feature extraction-based methods fetch features from epochs of data; thus, these features do not represent temporal information efficiently. Second, the large sample size might be the reason for the fast convergence and better performance of DeprNet. Since the data set is divided by splitting it into epochs of a 4-s sliding window with 75% overlapping, we get 17 307 samples without using any data augmentation technique. In comparison, other methods take input of either 7.8125 or 6 s without overlapping, which gives 2215 samples and 2885 samples, respectively. The third is the use of batch normalization. Batch normalization layers add stability and faster convergence of DeprNet.

TABLE V

ABLATION STUDY 1 ON SUBJECTWISE SPLIT DATA: THE CLASSIFICATION REPORT OF S-DEPRNET, T-DEPRNET, AND DEPRNET

	Precision	Recall	F1-score	Accuracy
S-DeprNet	0.479	0.533	0.505	0.537
T-DeprNet	0.869	0.822	0.845	0.860
DeprNet	0.919	0.887	0.895	**0.914**

TABLE VI

ABLATION STUDY 2 ON SUBJECTWISE SPLIT DATA: THE CLASSIFICATION REPORT OF DEPRNET WITH VARIABLE SAMPLE SIZE

Number of samples	Precision	Recall	F1-score	Accuracy
2000	0.669	0.617	0.642	0.670
10000	0.779	0.817	0.797	0.816
17307	0.919	0.887	0.895	**0.914**

TABLE VII

ABLATION STUDY 3 ON SUBJECTWISE SPLIT DATA: THE CLASSIFICATION REPORT OF DEPRNET WITH AND WITHOUT BATCH NORMALIZATION LAYERS

Batch normalization	Precision	Recall	F1-score	Accuracy
No	0.583	0.636	0.610	0.637
Yes	0.919	0.887	0.895	**0.914**

F. Ablation Study and Time Efficiency

The architecture of DeprNet is decided after performing ablation studies. In the first study, the effect of spatial dimension and temporal dimension on the performance of DeprNet is evaluated. S-DeprNet uses only spatial dimensions, T-DeprNet considers only temporal dimensions, and DeprNet utilizes both spatial and temporal dimensions. The sizes of the input layer of S-DeprNet, T-DeprNet, and DeprNet are 19, 1024, and 19×1024, respectively. Their performances are mentioned in Table V. In the second study, the effect of varying sample size on the performance of DeprNet is assessed. The results of this study are presented in Table VI. In the third study, the effect of the batch normalization layer on the performance of DeprNet is evaluated. The results of the third study are presented in Table VII. In the fourth study, the number of convolutional layers and batch normalization layers on the performance of DeprNet is assessed. The results of this study are presented in Table VIII.

TABLE VIII

ABLATION STUDY 4 ON SUBJECTWISE SPLIT DATA: THE CLASSIFICATION REPORT OF DEPRNET WITH VARIABLE NUMBER OF CONVOLUTIONAL LAYERS AND BATCH NORMALIZATION LAYERS

Number of layers	Precision	Recall	F1-score	Accuracy
0	0.502	0.552	0.526	0.558
1	0.483	0.545	0.516	0.540
2	0.739	0.703	0.720	0.746
3	0.848	0.886	0.867	0.879
4	0.911	0.888	0.899	0.910
5	0.919	0.887	0.895	**0.914**

TABLE IX

TIME EFFICIENCY (ONEFOLD) OF ACHCNN, ACHLSTM, T-LSTM, H-KNN1, H-KNN2, S-EMD, S-SVM, H-DBN, AND DEPRNET

Method	Training time	Testing time of one sample
AchCNN [23]	4859	0.0052
AchLSTM [24]	8652	0.0425
T-LSTM [27]	20254	0.2303
H-KNN1 [19]	36956	13.0665
H-KNN2 [20]	24852	6.5622
S-EMD [21]	18920	5.6340
S-SVM [46]	21888	15.6515
H-DBN [14]	29604	9.6442
DeprNet	**2230**	**0.0036**

The time efficiency of a NN is highly dependent on the parameters of a model and the specifications of the system's hardware. The training of DeprNet can be considered as two processes: data acquisition and model training. As suggested by Table VII, for training DeprNet to obtain 90%+ accuracy, around 17 307 samples (33 subjects) are needed. Each data acquisition session takes around 39 min, 20 min for setting up the EEG device and placing the electrodes on the scalp, 9 min for recording the EEG signals, and 10 min for removing and cleaning the electrodes. The time needed for recording the EEG is $39 \times 33 = 1287$ min. The time required for preprocessing the data set and training the DeprNet (one-fold) is $30 + 88 \times 25 = 2230$ s. Here, preprocessing takes 30 s and training of DeprNet with epoch size of 25, and a batch size of 64 takes $88 \times 25 = 2200$ s. On the other hand, the process of screening a patient for depression using DeprNet takes very little time. The data acquisition takes around 31 min, 20 min for setting up the EEG device and placing the electrodes on the scalp, 1 min for recording the EEG signals, and 10 min for removing and cleaning the electrodes. Although DeprNet needs only a short window of four seconds of EEG signals for predicting depression level, its performance can be improved by taking the average of predicted values of more than one sample; 1 min of recording the EEG signals will result in 15 samples (without overlapping), which can give better screening accuracy. The time required for preprocessing the data set and testing the DeprNet is $0.001 + 0.0036 \times 15 = 0.055$ s. Here, preprocessing takes 0.001 s, and forward computations of a model for one sample take 0.0036 s. The time efficiency of the abovementioned eight baseline methods and DeprNet is reported in Table IX. All abovementioned training and testing related timings are calculated on a computer system having specifications listed in Table II.

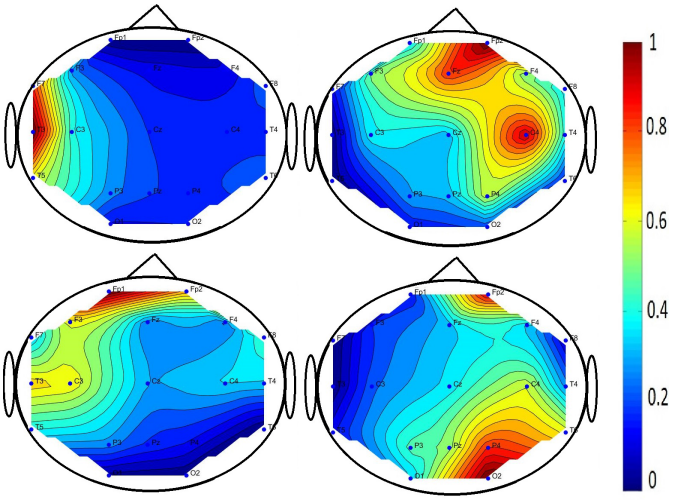

Fig. 11. **Top left**: sample topo heat map of values of the last layer of CNN of a nondepressed subject. **Top right:** sample topo heat map of values of the last layer of CNN of a depressed subject. **Bottom left:** average topo heat map of values of the last layer of CNN of a nondepressed subject. **Bottom right:** average topo heat map of values of the last layer of CNN of a depressed subject.

G. Analysis of the Features Learned by the Convolutional Layer

After analyzing the features learned by the last convolutional layer, it is found that most of the nondepressed subjects have higher values of the left-hand side electrodes than the value of right-side electrodes, while most of the depressed subjects have higher values of the right-side electrodes than the value of left-hand side electrodes. For testing the significance of visualization results, one-way Analysis of Variance (ANOVA) is performed on α values of depressed and nondepressed groups. A p-value of 0.0015 is obtained from ANOVA, which indicates that α is negative for a nondepressed subject and positive for a depressed subject. For graphical visualization of the results, we normalize the vector between 0 and 1 and show their values through a topo heat map for each subject. The sample topo heat map of a subject not having depression and the sample topo heat map of a subject having depression are shown in the upper half of Fig. 11. The averaged topo heat map of the nondepressed subject and the averaged topo heat map of the depressed subject are shown in the lower half of Fig. 11. These results suggest that depression affects the activities of different hemispheres of the brain differently. Moreover, this asymmetric effect is captured and used by CNN very efficiently for classification.

H. Discussion

The EEG signals are rich in information; they contain neural responses to the computations performed by the brain. Also, the greater affordability and portability of the EEG devices opened up new avenues for their use [48]. This study attempts to present a robust system that can be used for screening people for depression in practical scenarios. The psychiatrist can perform EEG tests on the patient regularly and keep track of their mental state during the treatment. However, there

are a few limitations of the proposed methods. The use of DeprNet with present technologies may be limited to clinical use only due to three reasons. First, it involves the use of an EEG device, which is a costly apparatus. Second, the data acquisition of EEG is a time-consuming process, and it also demands expertise. Third, DeprNet needs 19 EEG channels for classification, while the present personal use EEG headset mostly has only one electrode. Another limitation of DeprNet is the assumption that depression affects the patient's brain significantly. Minor changes in the patient's brain due to depression may not always be reflected in EEG signals. Thus, the use of EEG, which is an indirect measure of depression, may not always give accurate results.

The detection of depression using conventional methods, such as questionnaires, is a tricky task. Since these methods try to measure the severity of depression using indirect means, they sometimes fail to detect the exact level of depression. On the other hand, advanced technologies, such as EEG records, direct brain activity and can provide more robust results. However, the detection of mental disorders, such as depression, autism, and mood disorders from EEG signals, is not a trivial task. The EEG device indeed records neural activities related to these disorders, but dissociating these signals from noise is complicated due to their low amplitude. Modern advancements in DL methods can provide a solution to this problem as they can extract complex nonlinear features from the data.

We aim to develop two brain–computer interfaces (BCIs): one for the clinical use of psychiatric and another for the personal use of patients. The first BCI will use the EEG signal of the whole brain for detecting the level of depression, level of anxiety, and other abnormalities in the patient's brain. The BCI will upload the data to a cloud server and suggest possible treatments. The interface will serve as an assistant to the psychiatrists, and we help them in maintaining the history of the treatments of the patients. The second BCI will use EEG recording from a single electrode portable EEG headset. It will show a graphical representation of the report of the level of depression, level of anxiety, and other abnormalities in the patient's brain on a mobile phone app. This BCI will help them in regularly monitoring and improving their mental health. The method proposed in this study is an initial attempt in the direction of improving the mental health of individuals by providing better diagnosis techniques.

V. Conclusion

This study successfully makes use of DL models for analyzing the EEG data and demonstrating the transformation of brain activities in depression. It can be concluded that the DeprNet, a CNN-based DL model proposed in this study, performs better than the other baseline methods. Accuracy of 0.9937 and the AUC of 0.999 are achieved when recordwise split data are considered. Accuracy of 0.914 and the AUC of 0.956 are obtained, while subjectwise split data are adopted. These results suggest that CNN trained on recordwise split data gets overtrained on EEG data with a small number of subjects. However, most of the previous studies presented in Section II employed recordwise split data for the training and testing their models. It is also observed that the network can distinguish both the normal and depressed classes at the DeprNet level itself. The activation maps of the last layer of DeprNet suggest that the value of left electrodes is more than the values of right electrodes in nondepressed subjects, and the value of right electrodes is more than the values of left electrodes in depressed subjects. Moreover, the authors believe that depression affects the activities of both the hemispheres of the brain differently. The results obtained in this study are very promising, and this work can be extended by considering multiple factors in the future. Furthermore, based on the proposed diagnosis pipeline, a personalized mobile phone application can be developed to show the real-time depression level of a patient.

References

[1] A. Swetaa, R. Gayathri, and V. V. Priya, "Awareness of mental health among teenagers," *Drug Invention Today*, vol. 11, no. 8, pp. 1979–1982, 2019.

[2] M. A. Bell and K. Cuevas, "Using EEG to study cognitive development: Issues and practices," *J. Cognition Develop.*, vol. 13, no. 3, pp. 281–294, Jul. 2012.

[3] H. W. Cole and W. J. Ray, "EEG correlates of emotional tasks related to attentional demands," *Int. J. Psychophysiol.*, vol. 3, no. 1, pp. 33–41, Jul. 1985.

[4] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[5] S. In Cho, "Vision-based people counter using CNN-based event classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 8, pp. 5308–5315, Aug. 2020.

[6] H. Wang, S. Li, L. Song, L. Cui, and P. Wang, "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2648–2657, Jun. 2020.

[7] Z. Wan, J. Huang, H. Zhang, H. Zhou, J. Yang, and N. Zhong, "HybridEEGNet: A convolutional neural network for EEG feature learning and depression discrimination," *IEEE Access*, vol. 8, pp. 30332–30342, 2020.

[8] E. Bochinski, T. Senst, and T. Sikora, "Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3924–3928.

[9] S. D. Puthankattil and P. K. Joseph, "Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy," *J. Mech. Med. Biol.*, vol. 12, no. 4, Sep. 2012, Art. no. 1240019.

[10] M. Ahmadlou, H. Adeli, and A. Adeli, "Fractality analysis of frontal brain in major depressive disorder," *Int. J. Psychophysiol.*, vol. 85, no. 2, pp. 206–211, Aug. 2012.

[11] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, Mar. 2013.

[12] O. Faust, P. C. A. Ang, S. D. Puthankattil, and P. K. Joseph, "Depression diagnosis support system based on EEG signal entropies," *J. Mech. Med. Biol.*, vol. 14, no. 03, Jun. 2014, Art. no. 1450035.

[13] U. R. Acharya *et al.*, "A novel depression diagnosis index using nonlinear features in EEG signals," *Eur. Neurol.*, vol. 74, nos. 1–2, pp. 79–83, 2015.

[14] H. Cai, X. Sha, X. Han, S. Wei, and B. Hu, "Pervasive EEG diagnosis of depression using deep belief network with three-electrodes EEG collector," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1239–1246.

[15] W. Mumtaz, L. Xia, S. S. A. Ali, M. A. M. Yasin, M. Hussain, and A. S. Malik, "Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD)," *Biomed. Signal Process. Control*, vol. 31, pp. 108–115, Jan. 2017.

[16] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, Jun. 2017.

[17] G. M. Bairy *et al.*, "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features," *J. Med. Imag. Health Informat.*, vol. 7, no. 8, pp. 1857–1862, Dec. 2017.

[18] M. Sharma, P. V. Achuth, D. Deb, S. D. Puthankattil, and U. R. Acharya, "An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals," *Cognit. Syst. Res.*, vol. 52, pp. 508–520, Dec. 2018.

[19] H. Cai *et al.*, "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, 2018.

[20] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal EEG data for depression recognition," *Inf. Fusion*, vol. 59, pp. 127–138, Jul. 2020.

[21] J. Shen, X. Zhang, B. Hu, G. Wang, Z. Ding, and B. Hu, "An improved empirical mode decomposition of electroencephalogram signals for depression detection," *IEEE Trans. Affect. Comput.*, early access, Aug. 14, 2019, doi: 10.1109/TAFFC.2019.2934412.

[22] R. M. Sapolsky, "Depression, antidepressants, and the shrinking hippocampus," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 22, pp. 12320–12322, Oct. 2001.

[23] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 161, pp. 103–113, Jul. 2018.

[24] B. Ay *et al.*, "Automated depression detection using deep representation and sequence learning with EEG signals," *J. Med. Syst.*, vol. 43, no. 7, p. 205, Jul. 2019.

[25] X. Li *et al.*, "EEG-based mild depression recognition using convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 57, no. 6, pp. 1341–1352, Jun. 2019.

[26] D. Shah, G. Y. Wang, M. Doborjeh, Z. Doborjeh, and N. Kasabov, "Deep learning of eeg data in the neucube brain-inspired spiking neural network architecture for a better understanding of depression," in *Proc. Int. Conf. Neural Inf. Process.* Sydney, NSW, Australia: Springer, 2019, pp. 195–206.

[27] P. P. Thoduparambil, A. Dominic, and S. M. Varghese, "EEG-based deep learning model for the automatic detection of clinical depression," *Phys. Eng. Sci. Med.*, vol. 43, pp. 1349–1360, Oct. 2020.

[28] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, 1987.

[29] A. Seal *et al.*, "An EEG database and its initial benchmark emotion classification performance," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–14, Aug. 2020.

[30] R. Mahajan and B. I. Morshed, "Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, kurtosis, and wavelet-ICA," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 158–165, Jan. 2015.

[31] A. K. Maddirala and R. A. Shaik, "Separation of sources from single-channel EEG signals using independent component analysis," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 382–393, Feb. 2018.

[32] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: http://arxiv.org/abs/1505.00853

[33] A. Karlekar and A. Seal, "SoyNet: Soybean leaf diseases classification," *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105342.

[34] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[36] G. Tolias, R. Sicre, and H. Jégou, "Particular object retrieval with integral max-pooling of CNN activations," 2015, *arXiv:1511.05879*. [Online]. Available: http://arxiv.org/abs/1511.05879

[37] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. ICML*, 2016, vol. 2, no. 3, p. 7.

[38] S. Bruch, X. Wang, M. Bendersky, and M. Najork, "An analysis of the softmax cross entropy loss for learning-to-rank with binary relevance," in *Proc. ACM SIGIR Int. Conf. Theory Inf. Retr.*, Sep. 2019, pp. 75–78.

[39] M. Hänggi, S. Moser, E. Pfaffhauser, and G. S. Moschytz, "Simulation and visualization of CNN dynamics," *Int. J. Bifurcation Chaos*, vol. 09, no. 7, pp. 1237–1261, Jul. 1999.

[40] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[41] Z. Qin, F. Yu, C. Liu, and X. Chen, "How convolutional neural network see the world–a survey of convolutional neural network visualization methods," 2018, *arXiv:1804.11191*. [Online]. Available: http://arxiv.org/abs/1804.11191

[42] K. K. Sharma and A. Seal, "Clustering analysis using an adaptive fused distance," *Eng. Appl. Artif. Intell.*, vol. 96, Nov. 2020, Art. no. 103928.

[43] K. K. Sharma and A. Seal, "Multi-view spectral clustering for uncertain objects," *Inf. Sci.*, vol. 547, pp. 723–745, Feb. 2021.

[44] K. K. Sharma and A. Seal, "Modeling uncertain data using Monte Carlo integration method for clustering," *Expert Syst. Appl.*, vol. 137, pp. 100–116, Dec. 2019.

[45] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.

[46] M. Sharma, S. Patel, and U. R. Acharya, "Automated detection of abnormal EEG signals using localized wavelet filter banks," *Pattern Recognit. Lett.*, vol. 133, pp. 188–194, May 2020.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[48] L. Angrisani, P. Arpaia, A. Esposito, and N. Moccaldi, "A wearable brain–computer interface instrument for augmented reality-based inspection in industry 4.0," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1530–1539, Apr. 2020.

Ayan Seal (Senior Member, IEEE) received the Ph.D. degree in engineering from Jadavpur University, Kolkata, India, in 2014.

He has visited the Universidad Politecnica de Madrid, Madrid, Spain, as a Visiting Research Scholar. He is currently an Assistant Professor with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India. He has authored or coauthored several journals, conferences, and book chapters in the area of biometric and medical image processing. His current research interests include image processing and pattern recognition.

Dr. Seal was a recipient of several awards. Recently, he has received Sir Visvesvaraya Young Faculty Research Fellowship from the Media Lab Asia, Ministry of Electronics and Information Technology, Government of India.

Rishabh Bajpai received the B.Tech. degree in mechanical engineering from the PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India, in 2020. He is currently pursuing the Ph.D. degree with the Centre for Biomedical Engineering, IIT Delhi, New Delhi, India.

He has been working in the areas of brain–computer interface, rehabilitation engineering, neuroscience, electroencephalography, and the development of biomedical instrumentation for applications specific to assistive devices for the disabled.

Jagriti Agnihotri received the B.Tech. degree in computer science and engineering from the PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India, in 2020.

Her research interests include brain–computer interfaces, mental health systems, neural networks, and the application of machine learning and deep learning (DL).

Anis Yazidi (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Agder, Grimstad, Norway, in 2008 and 2012, respectively.

He was a Researcher with Teknova AS, Grimstad. From 2014 to 2019, he was an Associate Professor with the Department of Computer Science, Oslo Metropolitan University, Oslo, Norway, where he is currently a Full Professor and leading the research group in applied artificial intelligence. He is also Professor II with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His current research interests include machine learning, learning automata, stochastic optimization, and autonomous computing.

Enrique Herrera-Viedma (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 1993 and 1996, respectively.

He is currently a Professor of Computer Science and the Vice-President of Research and Knowledge Transfer with the University of Granada. His H-index is 85 with more than 25000 citations received in Web of Science and 97 in Google Scholar with more than 38500 citations received. His current research interests include group decision making, consensus models, linguistic modeling, aggregation of information, information retrieval, bibliometric, digital libraries, Web quality evaluation, recommender systems, and social media. He has been identified as one of the World's Most Influential Researchers by Shanghai Center and Thomson Reuters/Clarivate Analytics in both the computer science and engineering scientific categories in 2014–2020.

Prof. Herrera-Viedma was the 2019–2020 Vice-President of Publications with the IEEE SMC Society and an Associate Editor of several journals, such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, *Information Sciences*, *Applied Soft Computing*, *Soft Computing*, *Fuzzy Optimization and Decision Making*, and *Knowledge-Based Systems*.

Ondrej Krejcar received the Ph.D. degree in technical cybernetics from the Technical University of Ostrava, Ostrava, Czech Republic, in 2008.

He is currently a Full Professor of systems engineering and informatics with the University of Hradec Králové (UHK), Hradec Králové, Czech Republic, where he has been a Vice-Rector for science and creative activities since June 2020. He is also the Director of the Center for Basic and Applied Research, UHK. From 2016 to 2020, he was the Vice-Dean for Science and Research at the Faculty of Informatics and Management, UHK. At UHK, he is a Guarantee of the Doctoral Study Program in Applied Informatics, where he is focusing on lecturing on Smart Approaches to the Development of Information Systems and Applications in Ubiquitous Computing Environments. His H-index is 18, with more than 1150 citations received in the Web of Science. His research interests include control systems, smart sensors, ubiquitous computing, manufacturing, wireless technology, portable devices, biomedicine, image segmentation and recognition, biometrics, technical cybernetics, and ubiquitous computing. His second area of interest is in biomedicine (image analysis), as well as biotelemetric system architecture (portable device architecture and wireless biosensors), and the development of applications for mobile devices with the use of remote or embedded biomedical sensors.

Dr. Krejcar has been a Vice-Leader and a Management Committee Member at WG4 of the project COST CA17136 since 2018. He has been a Management Committee Member Substitute for the project COST CA16226 since 2017. In 2018, he was the 14th top peer reviewer in Multidisciplinary in the World according to Publons and a Top Reviewer in the Global Peer Review Awards 2019 by Publons. He is also on the Editorial Board of the *Sensors* (MDPI) IF journal (Q1/Q2 at JCR) and several other ESCI indexed journals. Since 2019, he has been the Chairman of the Program Committee of the KAPPA Program, Technological Agency of the Czech Republic. He is also a Regulator of the EEA/Norwegian Financial Mechanism in the Czech Republic for the term 2019–2024. Since 2020, he has been Chairman of the Panel 1 (Computer, Physical and Chemical Sciences) of the ZETA Program, Technological Agency of the Czech Republic. From 2014 to 2019, he was the Deputy Chairman of the Panel 7 (Processing Industry, Robotics, and Electrical Engineering) of the Epsilon Program, Technological Agency of the Czech Republic.