GENet: A Generic Neural Network for Detecting Various Neurological Disorders From EEG

Md. Nurul Ahad Tawhid , Siuly Siuly , *Member, IEEE*, Kate Wang , and Hua Wang , *Senior Member, IEEE*

Abstract—The global health burden of neurological disorders (NDs) is vast, and they are recognized as major causes of mortality and disability worldwide. Most existing NDs detection methods are disease-specific, which limits an algorithm's cross-disease applicability. A single diagnostic platform can save time and money over multiple diagnostic systems. There is currently no unified standard platform for diagnosing different types of NDs utilizing electroencephalogram (EEG) signal data. To address this issue, this study aims to develop a generic EEG neural Network (GENet) framework based on a convolutional neural network that can identify various NDs from EEG. The proposed framework consists of several parts: 1) preparing data using channel reduction, resampling, and segmentation for the GENet model; 2) designing and training the GENet model to carry out important features for the classification task; and 3) assessing the proposed model's performance using different signal segment lengths and several training batch sizes and also cross-validating using seven different EEG datasets of six distinct NDs namely schizophrenia, autism spectrum disorder, epilepsy, Parkinson's disease, mild cognitive impairment, and attention-deficit/hyperactivity disorder. In addition, this study also investigates whether the proposed GENet model can identify multiple NDs from EEG. The proposed model achieved much better performance for both binary and multiclass classification compared to state-of-the-art methods. In addition, the proposed model is validated using several ablation studies and layerwise feature visualization, which provide consistency and efficiency to the proposed model. The proposed GENet model will help technologists create standard software for detecting any of these NDs from EEG.

Index Terms—Attention-deficit/hyperactivity disorder, autism spectrum disorder (ASD), convolutional neural network (CNN), electroencephalogram (EEG), epilepsy, mild cognitive impairment, Parkinson's disease, schizophrenia.

I. INTRODUCTION

NEUROLOGICAL disorders (NDs) are a set of diseases that damage both the central and peripheral nervous systems and include everything from neurodegenerative to neurodevelopmental and psychiatric conditions [1]. There are more than 600 types of NDs in the world [2] including schizophrenia (SZ), Parkinson's disease (PD), mild cognitive impairment (MCI), epilepsy (EP), Alzheimer's disease (AD), and dementia [1]. Additionally, developmental disorders such as autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) are also classified as NDs [1]. According to the World Health Organization (WHO), ND is the second biggest cause of mortality and the primary source of disability [2]. Early detection and treatment can improve the health of ND patients; however, early diagnosis is difficult due to the lack of computer-aided diagnosis (CAD) systems and the shortage of mental health professionals.

Electroencephalogram (EEG) is the most extensively utilized brain activity-capturing technique due to its excellent temporal resolution, availability, noninvasiveness, economical set-up costs, and widespread availability for professionals [3], [4], [5]. EEG captures the electrical activities of neurons from the human brain as signal data, and those signals are then visually analyzed by the expert clinicians for the identification of NDs. This visual analysis process is subjective, lengthy, error-prone, and difficult due to the overlapping features for different diseases, which may lead to misdiagnosis [6]. Additionally, the availability of neurologically trained workers differs by a factor of 70 between high- and low-income countries (1 versus 70 for 100 000 people) [2]. As a result, the development of CAD systems will aid doctors and improve ND diagnosis at earlier stages.

The major challenge of the existing EEG signal classification system is its applicability across different NDs [7], [8], [9], [10], [11], [12], [13], [14], [15]. For example, a method proposed to deal with a specific EEG classification problem may not be totally efficient in its identical configurations for an EEG signal classification problem in another disease. This is due to the nonstationarity, nonlinearity, and strong localization in the temporal, spectral, and spatial dimensions of the EEG signal [16]. For every ND like SZ, PD, ADHD, MCI, EP, and ASD, the underlying characteristic of EEG has distinct periodical and statistical properties, making it challenging to examine with a detection method created for a different ND. Therefore, the

motivation of this research is to provide a unified process for EEG signal classification that is adaptable and effective for a wide range of EEG challenges.

With the advancement of technology, CAD has become an important part of the medical industry. Several studies have been conducted to diagnose NDs using EEG data. EEG signal classification techniques can be broadly classified into two categories based on the used feature extraction and classification approaches, namely machine learning (ML)-based classification and deep learning (DL)-based classification. In ML-based techniques, handcrafted features like statistical and nonlinear parameters are extracted from the time, frequency, and time–frequency domains, and different ML-based classifiers are used to perform categorization on those extracted features. ML-based classification of EEG signal is carried out in several studies for the identification of SZ [7], [17], [18], MCI [8], PD [9], ASD [19], and ADHD [20]. This process requires experts on those feature domains for the analysis of EEG data.

In DL-based classification, several DL models exist, like convolutional neural network (CNN), long short-term memory (LSTM), recurrent neural network (RNN), and gated recurrent unit (GRU), among which CNN is widely used for image and signal processing, natural language processing, and data analytics [21]. CNN is generally less sensitive to noise and can extract useful information from noisy input [22]. Although CNN has been proven extremely effective for image classification, it has also demonstrated efficiency in EEG signal categorization. Previously, CNN-based EEG data classification was completed for SZ [10], EP [11], [18], [23], PD [12], [18], [24], ASD [13], [18], [21], MCI [14], [25], brain–computer interfaces (BCI) [26], and ADHD [15] had good classification accuracy. However, those studies focused on a single disease identification which makes their multidisease applicability an issue. Moreover, most of the studies have converted the EEG signals to visual representations or extracted handcrafted features before classifying them using the CNN model. Due to this handcrafted feature extraction and signal visualization, the CNN model cannot learn the significant features that it can extract from the raw EEG data. This study addresses these gaps by developing a multidiseases scalable framework using raw EEG signal data as input for the CNN model.

In this study, we have developed a CNN model based on the Generic EEG neural Network (GENet) that can categorize EEG signals of different NDs. Furthermore, we have used the raw EEG data as an input to the GENet model so that the DL process can self-learn the significant features from the EEG signal. At first, the raw EEG data are preprocessed to make those signals ready for input into the GENet model. To do that, we have reduced the total number of channels to 19 channels according to the international 10–20 standard [27], [28] and resampled the signals to 256 Hz. Then, to use the short-term features of the EEG signal, we have segmented the signals into small time frames. Here, we have tested three different segment lengths to check the impact of the signal segment length on the detection process. Finally, the GENet model is trained using those signals and performed the classification task. Seven different EEG datasets from six different NDs have been

used to validate the proposed model. We have also performed a five-class classification using four datasets from among those seven datasets. A five-fold cross-validation technique is used to remove any biases in the results. Results from this study are compared with existing state-of-the-art studies that have used the same EEG datasets.

The major contributions of this study are compiled as follows.

1) A novel framework has been created to address the scalability challenge encountered in multidisease studies within existing research. This framework introduces a generic EEG classification approach using the CNN model to identify various NDs. Its scalability has been validated by assessing seven distinct datasets, each representing one of six major NDs.

2) The impact of diverse EEG signal segment lengths on the classification process is assessed by testing three different segment durations.

3) The presented research framework successfully resolves a substantial issue in addressing the multidisease classification problem within the EEG-based ND detection field, thereby bridging a significant gap in existing research.

4) The proposed computerized system is validated through extensive experiments including classifier performance evaluations, layerwise classification feature visualization, ablation studies, and comparative analyses with recent notable research.

5) The proposed GENet model was explored in both binary and multiclass classification scenarios, demonstrating superior performance compared to existing methodologies.

The rest of the article is structured as follows. Section II gives the details of the proposed method used in this study. Classification performance of the proposed model is described in Section III. Finally, Section IV gives the concluding remarks and future works of this study.

II. MATERIALS AND PROPOSED METHODS

In this section, the proposed framework, the datasets used to validate the framework, and the evaluation parameters are introduced in detail. In Section II-A, we have discussed details of the datasets that are used in this study followed by the different processing steps of the proposed framework in Section II-B. Proposed classification model is presented in Sections II-C and II-D and finally different evaluation parameters and approaches are discussed in Section II-E. Fig. 1 illustrates the schematic diagram of the proposed framework. A detailed discussion of those steps is given in the below subsections.

A. EEG Data Collection

In this research, we have used seven different publicly available datasets for six different NDs, namely: ASD, ADHD, EP, PD, MCI, and two datasets for SZ. We have chosen these datasets as those are widely used in different studies and are
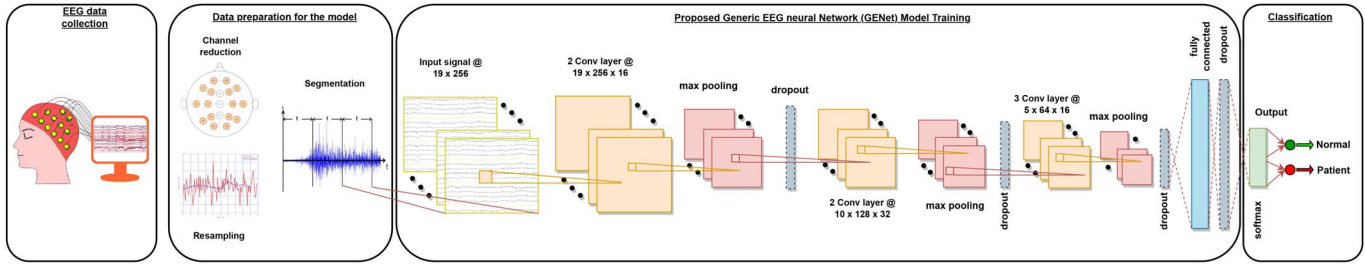
Fig. 1. Schematic illustration of the proposed framework and steps involved in the analysis. Four modules of the framework are discussed in four Sections II-A, II-B, II-C, and II-D.

publicly available. A brief description of those datasets is given as follows.

1) The first schizophrenia dataset (hereafter referred to as SZ1) that we have used in this study was collected from the Kaggle website [29]. It contains 81 subjects [49 SZ, 32 healthy control (HC)]. EEG data are recorded from 64 channels at a sampling rate of 1024 Hz during a task of press button tone.

2) Second schizophrenia dataset (hereafter referred to as SZ2) is comprised of 28 subjects (14 age- and sex-matched subjects from the SZ and HC groups) that were collected at the Institute of Psychiatry and Neurology in Warsaw, Poland [30]. The signals are recorded in their resting state at a sampling rate of 250 Hz from 19 channels of a standard 10–20 EEG electrode system.

3) Dataset of MCI is collected from cardiac catheterization units of Sina and Nour Hospitals, Isfahan, Iran [31]. This dataset consists of 27 subjects' resting state EEG signals (11 MCI and 16 HC) that were recorded from 19 electrodes at a sampling rate of 256 Hz.

4) ADHD dataset is recorded in the Psychology and Psychiatry Research Center at Roozbeh Hospital (Tehran, Iran). It has a total 121 subjects with 61 from ADHD group and 60 from control group. Visual attention task-based EEG recording is carried out from 19 channels using standard EEG 10-20 system at a sampling frequency of 128 Hz.

5) Epilepsy dataset is collected in Universidade Federal do Para, Brazil with 14 subjects (7 EP, 7 HC) [32]. EEG data is recorded in 256 Hz sampling rate from 20 channels while the subjects were in resting state.

6) For PD, a dataset from the University of Iowa, USA, is used here [33]. This dataset has a total of 28 subjects (14 age- and sex-matched subjects from each group). EEG signal data in resting state is recorded from 64 channels at a sampling rate of 500 Hz.

7) ASD dataset consists of 16 subjects (12 ASD, 4 non-ASD), which were collected from King Abdulaziz University (KAU) Hospital, Saudi Arabia [34]. Resting state EEG data is collected from 16 channels at a sampling frequency of 256 Hz.

Each dataset is chosen because of its distinctive underlying temporal and spectral characteristics, which are critical for the development of a uniform EEG categorization model. But due to the limited space availability, thorough descriptions of these
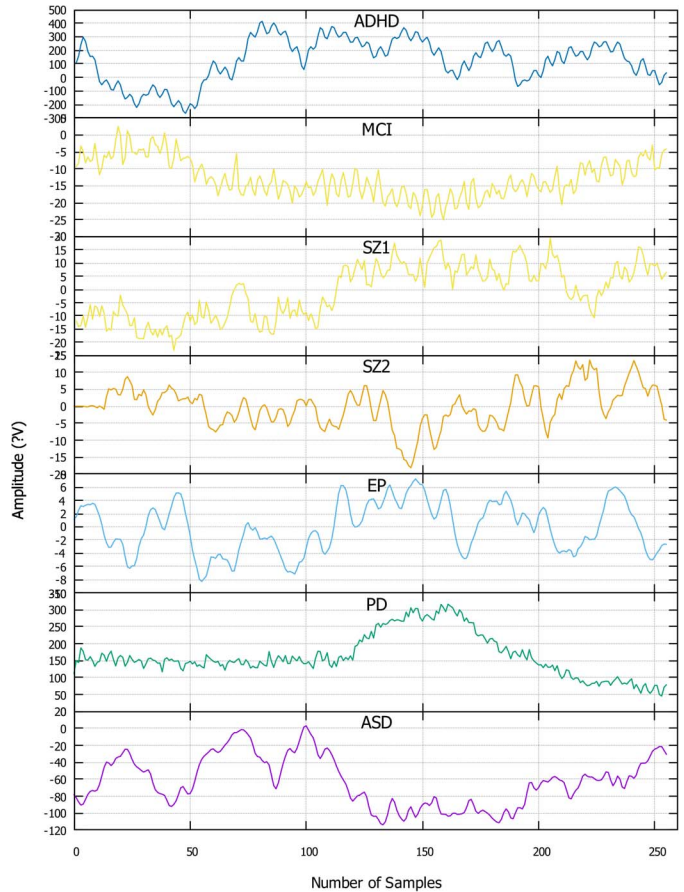
Fig. 2. 256 sample point data from the Fp1 channel of the tested seven datasets.

datasets have been excluded. Table I provides a summary of the participants' demographic data for different datasets. Details of those datasets can be found in [29], [30], [31], [32], [33], and [34]. Sample data points from the Fp1 channel of the tested seven datasets are given in Fig. 2.

Each participant provided their informed consent to the publication of their data at the time the data was gathered, and all of these datasets are openly available online. The confidentiality of the participants was also protected by not posting any personally identifiable information about the respondents; therefore, no ethical approval was required for our study.

TABLE I
DEMOGRAPHIC DATA FOR THE DATASETS USED IN THIS STUDY

	SZ1 [29]	SZ2 [30]	MCI [31]	ADHD [35]	EP [32]	PD [33]	ASD [34]
Patients (P) (Male/Female)	49 (41/8)	14 (7/7)	11	61 (48/13)	7 (4/3)	14 (6/8)	12 (9/3)
Normal (N) (Male/Female)	32 (26/6)	14 (7/7)	16	60 (50/10)	7 (4/3)	14 (6/8)	4 (4/0)
Patients age range	40.02 ± 13.7	28.1 ± 3.7	66.4 ± 4.6	9.62 ± 1.75	32.86 ± 9.51	70.5 ± 8.35	12.5 ± 3.91
HCs age range	38.38 ± 13.7	27.75 ± 3.15	65.3 ± 3.9	9.85 ± 1.77	32.86 ± 9.51	70.5 ± 8.35	11 ± 2.49
Sampling frequency	1024	256	256	128	256	500	256
Resampled frequency	256	256	256	256	256	256	256
Recorded no of channel	64	19	19	19	20	64	16
Used no of channel	19	19	19	19	19	18	16
Number of samples generated after signal segmentation							
1 second segment (P/N)	14184/9324	15457/12716	21009/29931	9404/7466	3744/3707	2606/2657	11486/4848
2 second segment (P/N)	7083/4655	7725/6356	10503/14960	4683/3717	1872/1853	1299/1326	5737/2421
3 second segment (P/N)	4728/3108	5146/4235	6999/9972	3117/2471	1248/1235	864/881	3825/1612

B. Preparing Data for the Proposed Model

In this step, we have completed some preprocessing of the EEG signals to make the data ready for input into the proposed model. Here, the preprocessing of the raw EEG signal consists of three subtasks: 1) picking the standard common channels to use in the classification process. 2) Resampling the signals. 3) Segmenting the signals into small time frame blocks. Details of those steps are discussed in the following sections.

1) Channel Reduction to Use Standard Channel Data: From Table I, we see that the number and position of the recording channels vary from dataset to dataset. To prepare those data for inputting into the proposed CNN model, we have selected the channels that are common to the datasets and also standard for EEG signal analysis. To do so, we have selected the 19 most widely used channels (Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3/T7, T4/T8, P3, P4, T5/P7, T6/P8, O1, O2, Fz, Cz, and Pz) from the international 10–20 system for their standardized placements, offering a balance between coverage and compatibility across studies [27], [28] as shown in Fig 3. For the PD dataset, although it has recorded 64 channels, it was missing the Pz channel in those 64 channels, and for the ASD dataset, the original dataset only contains 16 channels of EEG recording. As a result, we modified our proposed model to accept 18 and 16 channel data as input for the classification of PD and ASD, respectively.

2) Resampling the Signals to a Standard Frequency: After channel selection, the next thing we did was resample all the datasets to a common frequency band so that they could be input into the proposed model. From Table I, we can see that four datasets have 256 Hz sampling rate, which is a widely used sampling frequency for EEG data and computationally less expensive compared to high-frequency bands [36]. We have also chosen this frequency band as the standard for the proposed model and converted all other datasets into a 256 Hz sampling rate.

3) Segmentation of the EEG Signals: Data scarcity is a major issue in the field of EEG signal analysis using DL-based techniques. This issue is often solved by the researchers using segmentation techniques. In this process, original EEG data are
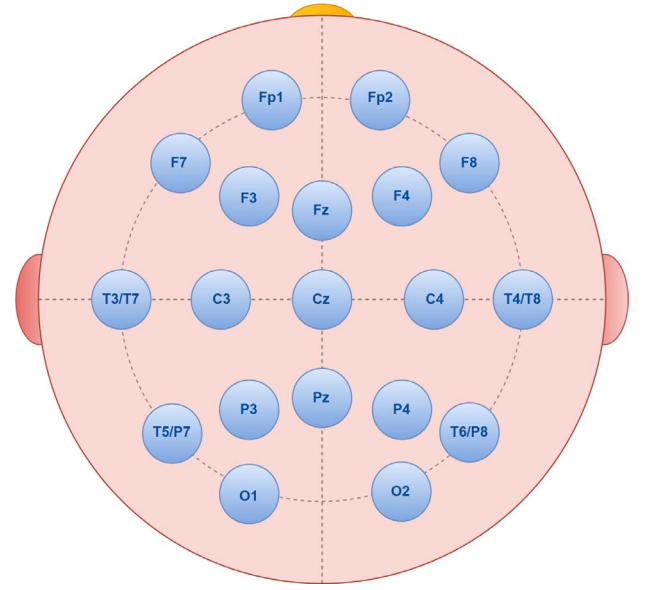
Fig. 3.   Standard electrode locations used for EEG data recording using the international 10–20 system.

segmented into small informative segments and are given the same level as the original one which results in an increase in data sample size with equal ratio [19], [37]. In this study, we have tested three different segment lengths (1, 2, and 3 s) to check the effect of segments on the classification process as well as the minimum length of the EEG signal, which is enough for representative feature extraction and disease classification. In the bottom three rows of Table I, we have given total number of samples generated after segmentation of 1, 2, and 3 s, respectively.

C. Proposed Generic EEG Neural Network (GENet) Model

In this study, we have developed a CNN model named GENet to perform the classification of the raw EEG signal data. We have used the CNN model as it is usually less sensitive to noise and can extract useful information from noisy input

Fig. 4. GENet: the proposed CNN model.

TABLE II
ARCHITECTURAL DETAILS OF THE GENET MODEL

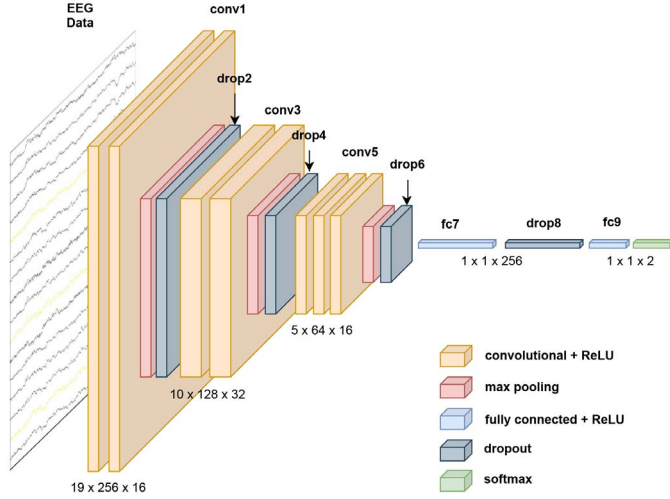| Layer	# Filter	Size	Activation	Option
Conv2D	16	3x3	ReLU	Padding=same
Conv2D	16	3x3	ReLU	Padding=same
MaxPooling2D		1x2		Stride=2x2
Dropout				25%
Conv2D	32	3x3	ReLU	Padding=same
Conv2D	32	3x3	ReLU	Padding=same
MaxPooling2D		2x2		Stride=2x2
Dropout				25%
Conv2D	16	3x3	ReLU	Padding=same
Conv2D	16	3x3	ReLU	Padding=same
Conv2D	16	3x3	ReLU	Padding=same
MaxPooling2D		2x2		Stride=2x2
Dropout				25%
Flatten				
Dense	256		ReLU	
Dropout				50%
Dense (classifier)	2 (binary)/ 5(multi)		softmax	

Total trainable params: 943 906

by learning suitable features on its own using convolutional kernels, filtering, pooling, and nonlinear activation operations, and classifying data into different categories [22]. Architectural diagram of the proposed GENet model is shown in Fig. 4.

The GENet model contains seven convolution (Conv2D) layers, three max-pooling (MaxPooling2D) layers, four dropout layers, and a fully connected (FC) layer. Table II lists details configuration of those layers. Conv2D layer consists of multiple kernels for feature extraction, and the local connection and weight sharing characteristics are used to reduce network parameters and overfitting. Conv2D layer operation can be defined using as follows, which involves multiplying input data with a convolutional kernel and adding an offset, with the kernels sequentially scanning the input data of the upper layer [38].

$$y(i,j) = \sum_{k=1}^{n} \sum_{l=1}^{m} \omega(k,l)x(i-k,j-l) + b \quad (1)$$

where $y(i,j)$ is the ith row and jth column output feature map, $w(k,l)$ is the filter weights, $x(i-k,j-l)$ is the input data, and b is the bias term. The first two Conv2D layers of the GENet model have 16 filters with a kernel size of (3×3) and a stride of one pixel. After each convolutional layer, the activation function is utilized to activate the node's summed input. Here, rectified linear unit (ReLU) is used, which has a linear identity for all positive values and assigns zero to all negative values using as follows on input x [38]. ReLU has the benefit of good generalization with less computational cost.

$$R(x) = \max(0, x) \quad (2)$$

Following those two Conv2D layers is a max-pooling layer with a pool size of (1×2) and a stride of (2×2). Pooling layers are used to downsample the output from the convolution layer to reduce the dimensionality of the feature maps and introduce some degree of translation invariance [38]. The most common pooling operation is max pooling, which selects the maximum value from a local neighborhood using as follows:

$$y_{i,j,k} = \max_{m,n} x_{i \times s+m, j \times s+n, k} \quad (3)$$

where $y_{i,j,k}$ is the output of the kth feature map at position (i,j), $x_{i \times s+m, j \times s+n, k}$ is the input at position $(i \times s + m, j \times s + n)$ and the kth feature map, and s is the stride.

Following the pooling layer, we have used a 25% dropout layer to regularize the CNN model and avoid overfitting issues. At each training epoch, a random portion of the layer's nodes are dropped out (set to zero) via Dropout. As a result, the risk of overfitting is reduced, and the network is forced to learn more robust features.

The third and fourth Conv2D layers have 32 filters with (3×3) kernels and one pixel stride. These two Conv2D layers are also followed by a MaxPooling2D layer with pool size and stride, both of which are (2×2), and a 25% dropout layer. The fifth, sixth, and seventh Conv2D layers have 16 filters each with a kernel of (3×3) and one pixel stride. Then, there is a MaxPooling2D layer with a pool and stride of (2×2), followed by a dropout layer of 25%.

FC layer connects all the neurons from the previous layer to the output layer. It accepts flattened output from the convolutional layers and performs a linear transformation of the input followed by an activation function [38]. The internal equation of FC layer can be denoted as follows:

$$y_k = \sigma\left(\sum_i w_{i,k} x_i + b_k\right) \quad (4)$$

where y_k is the kth neuron's output, σ is the activation function, $w_{i,k}$ is the weight parameter connecting the ith input neuron to the kth output neuron, x_i is the ith neuron's input, and b_k is the bias parameter for the kth output neuron. We have used an FC layer with 256 neurons that is activated by ReLU functions, followed by a 50% dropout layer. Finally, the classification layer uses 2 or 5 neurons, based on the number of classes for classification (binary or multi). The softmax activation function is used for this layer, which uses in the following

equation for internal calculation.

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (5)$$

where y_i is the ith element of the output vector, x_i is the ith element of the input vector, and n is the size of the input vector.

We have used categorical cross-entropy as the loss function that measures the cross-entropy loss between the labels and predictions and is used in classification problems with two or more label classes. Previously, several studies have used this loss function in EEG signal classification and obtained good classification results [18], [21], [39], [40], [41]. To minimize the loss function, the Adam algorithm, a stochastic gradient descent (SGD) technique based on the adaptive learning rate of the 1st- and 2nd-order moments of the gradient average, is used as the optimizer. This approach typically accelerates the model's convergence and is more resistant to noise and sparse gradients.

D. Classification Using Proposed GENet Model

After feature extraction and GENet model training, the classification process is carried out on the test cases in the final dense layer. For binary classification, this layer contains two neurons, while for multiclass classification, it has five neurons, as we have conducted a five-class classification process by modifying the last layer. To evaluate the performance of the GENet model in binary classification, we have tested it on seven different EEG datasets from six distinct NDs shown in Table I. Furthermore, to assess the performance of the proposed framework on multiclass classification, we have performed a multidisease detection using four of the tested datasets (SZ2, MCI, EP, and ADHD) from Table I. In this case, we have merged all the HC subjects from those four datasets into one group (HC) and performed a five-class classification process (SZ versus MCI versus EP versus ADHD versus HC).

E. Performance Evaluation Criteria

The number of correct classifications may depend on the training set and the test set, so cross-validation is one way to identify a model's prediction accuracy and decrease over-fitting [21]. A technique to achieve this is known as the k-fold cross-validation, which starts by randomly dividing the dataset into k subsets of equal or nearly equal size, with k-1 subsets used for training and the rest used for testing. This training/testing process is repeated k times (k-fold), using a different subset for testing every time. Here, we have evaluated the models' performance using a 5-fold cross-validation technique.

We have also used the leave-one-out validation (LOOV) where we have excluded all data from a single subject and utilized information from all remaining subjects to train the model. The excluded subject's data were then employed for validation. This iterative process is repeated for each subject, and the ultimate performance metric was derived from averaging the results obtained from all subjects.

Finally, to report the performance of the proposed framework, we have used five well known evaluation parameters in

this field which are sensitivity (Sen) or recall, specificity (Spec) or selectivity, precision (Prec), F1 score (F1) and accuracy (Acc). We have also considered the receiver operating characteristic (ROC) curve for validation purposes. These criteria can be used to anticipate how classifiers would behave on test data [7], [8], [39], [41], [42], [43].

III. RESULTS AND DISCUSSION

In this research, a generic framework is proposed for the classification of EEG data from patients and healthy subjects. To validate the proposed model's scalability, we have tested it on seven distinct EEG datasets from six different NDs. The next subsection discusses the detailed experimental setup for the proposed system, and the detailed results of the experiments are discussed in the later subsections.

A. Experimental Setup

As stated previously in the methodology section, after preprocessing, we have segmented the EEG signals into small time frames. To check the impact of the segment length on the proposed model, we have tested three different segment lengths: 1, 2, and 3 s. After segmentation, an EEG signal trial form a matrix of $s \times c$, where s is the signal length and c is the number of channels. s can be defined by $l \times f$, where l is the segment length (1, 2, or 3 s) and f is the sampling frequency (256 Hz). Therefore, the CNN input matrix sizes for 1, 2, and 3 s segment lengths are 256×19, 512×19, and 768×19, respectively.

Afterword resulted signal populations are arbitrarily divided into five equal or nearly equal subparts to carry out the 5-fold cross-validation for all the datasets. In this cross-validation scheme, four out of five sub parts are used to train the proposed model and the rest one is used to validate it. This approach is repeated five times to verify that each segment is used exactly once for testing the models. On the other hand, for LOOV, all the signal segments of a subject are left out from the training process and used for testing the trained model. This process is repeated for total number of subjects in the dataset.

The experiments are performed on a computer equipped with an Intel(R) Core(TM) i5 CPU @1.7GHz processor, 8 GB memory, Windows 10 64-bit operating system, and Google Colab[1]. The GENet model is trained with 50 epochs, as it starts overfitting after that. Training batch size selection is done using mini-batch mode, a popular batch size selection approach for faster learning [21]. We have tested four different batch sizes (32, 64, 128, and 256) during the training process of the proposed model. All the training processes on the proposed model are carried out from the scratch.

B. Layerwise Feature Visualization of GENet

T-distributed stochastic neighbor embedding (t-SNE) is a nonlinear dimension reduction approach that projects multivariate data on a 2-D or 3-D space in an unsupervised manner [44]. We have used t-SNE visualization to generate 2-D representations of the extracted features from each layer of the GENet

---

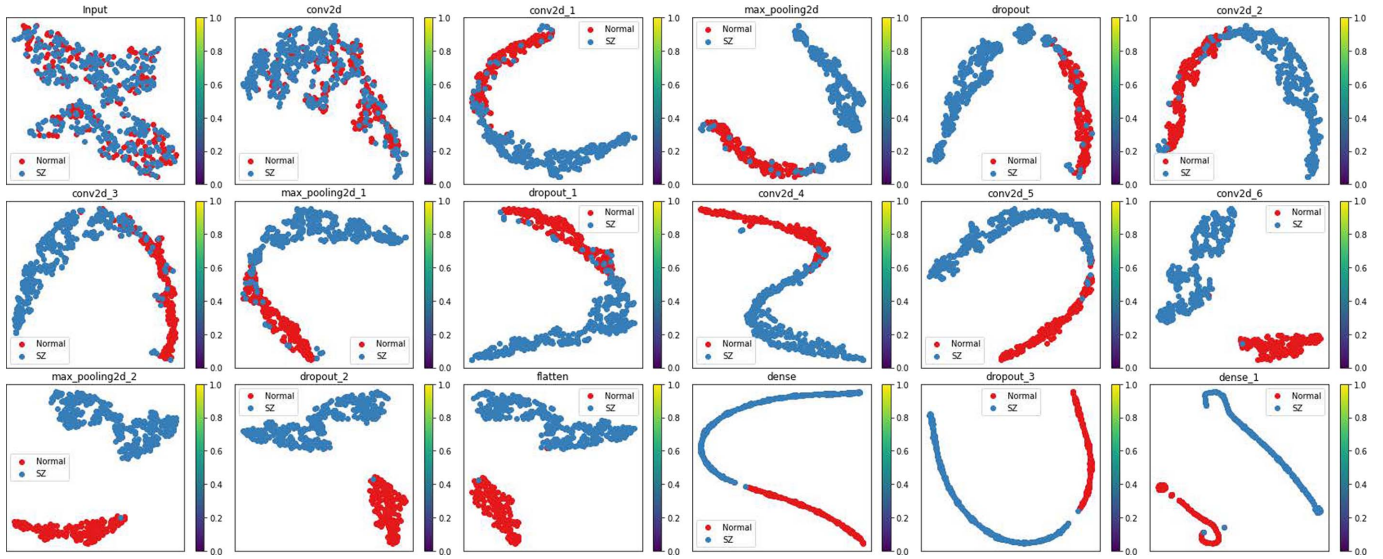[1]https://colab.research.google.com/notebooks/intro.ipynb

Fig. 5. Layerwise classification process visualization of the GENet model using t-SNE images. Here, features from test subjects are plotted from input layer to the output layer for tested SZ2 dataset. At the input layer, there was no clear cluster between two classes (SZ versus Normal), but as the data progressed from the hidden layers to the output layer, they formed two clearly separable clusters of two classes.

model. This technique helps to visualize the model's layerwise extracted features in the classification process. Fig. 5 illustrates the feature visualization from input to output layer of the GENet model for the tested SZ2 dataset. For the sake of convenience, we have included a single fold's t-SNE plots with 300 test subjects. The figures show the 2-D map of the multidimensional feature vectors, with each symbol representing an individual sample from the test set.

t-SNE plot is a useful way to visualize how the extracted features of a classifier from different categories are clustered together and are well separable or not [44]. From Fig. 5, we can see that in the input layer projection, all of the feature points from the two classes are randomly mixed up, and as the data passes through the layers of the GENet model, it is clustered into two classes. Finally, in the t-SNE plot of the output layer, we can see the completely separable clusters of the two groups (SZ versus HC). Hence, from these t-SNE images, we can claim that our proposed GENet model performs well on EEG data analysis for anomaly detection.

C. Results

We have performed two types of classification tasks, binary, and multiclass classification. The detailed results of these two experiments are discussed as follows.

1) Binary Classification: For binary classification, we have tested the performance of the proposed model on seven different datasets from six distinct NDs. To check the impact of segment length, three different segment lengths are tested: 1, 2, and 3 s, and those results are compared. Furthermore, to test the impact of the training batch size on the GENet model, we have used four separate batch sizes (32, 64, 128, and 256) to train the model. Fig. 6 plots the 5-fold average accuracy comparison for three segment lengths and four batch sizes on the tested

datasets. From the figure, we can see that for most of the cases, a segment length of 1 s produces the best classification accuracy compared to the other two segment lengths. For this reason, further discussions of this study are based on 1 s segment length.

To further evaluate the performance of the GENet, we have used four evaluation parameters that are popular in this field of study: sensitivity, specificity, precision, and F1 score. An experiment's sensitivity (also known as recall, hit rate, or true positive rate) refers to the classifier's ability to correctly distinguish patients from healthy people. On the other hand, the capacity of a test to correctly separate healthy participants from patients is referred to as specificity (also known as true negative rate or selectivity). Precision, also known as "positive predictive value" in the classification context, refers to the percentage of true patients in the retrieved patient group. Finally, F1 score is calculated by combining the precision and recall values of a test result to determine the test's performance in finding patients. It is the harmonic mean of precision and recall. The performance of a classification framework expects high values for those parameters.

Table III lists the five-fold average values with standard deviation (SD) of the performance parameters for the tested datasets with four different training batch sizes. From Table III, we can see that for most of the datasets, batch size variation has not that much impact on the model's performance except for PD and ASD, where parameter values decrease with the increase of training batch size. For the tested dataset SZ 1, the highest sensitivity value of 99.10($\pm$0.31)% is achieved for batch size 128 while the best specificity and precision values of 98.94($\pm$0.24)% and 99.30($\pm$0.15)% (pm 0.15) are obtained for training batch size 64. Five-fold average highest accuracy 98.82($\pm$0.29)% is achieved using the training batch size 128. In the case of the F1 score, all the tested training sizes produce the same value of 0.99. For dataset
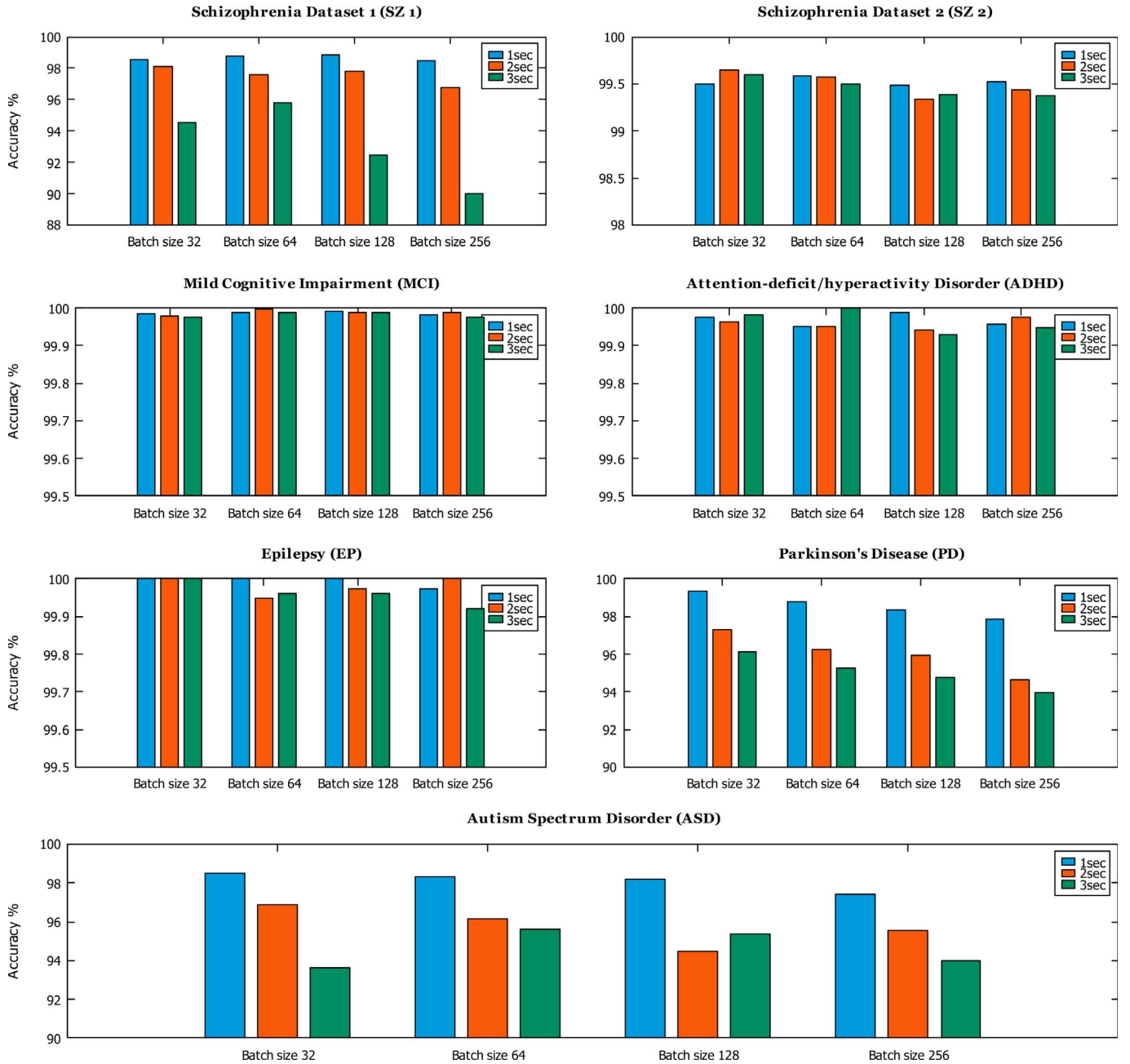
Fig. 6.   Accuracy comparison of the GENet model for the three tested signal segment lengths (1, 2, and 3 s). Seven subplots represent seven tested datasets (SZ1, SZ2, MCI, ADHD, EP, PD, and ASD). In each subplot, four groups of bars represent four training batch sizes (32, 64, 128, and 256).

SZ 2, the highest values for the five evaluation parameters are 99.65(±0.11)%, 99.54(±0.12)%, 99.62(±0.1)%, 1.00(±0.00), and 99.58(±0.11)% for the training batch sizes of 64, 128, 128, 256, and 64, respectively.

In MCI and ADHD datasets, our proposed model has achieved five-fold highest accuracy of 99.99% with SD of (±0.01) and (±0.02), respectively. It has obtained 100(±0.00)% sensitivity and 99.99% specificity and precision with SD (±0.02) and (±0.01), respectively, in the classification performance for MCI dataset. On the other hand, for the ADHD dataset, sensitivity is 99.98(±0.03)% and specificity and precision values are 100(±0.00)%. For both the datasets, F1

scores are 1.00. Among the tested datasets, for the EP dataset, we have achieved an overall 100(±0.00)% classification accuracy for three batch sizes (32, 64, and 128). For batch size 256, the performance of the model has decreased to 99.97(±0.04)%.

As mentioned in Table I, both the PD and ASD datasets have a smaller number of channels than other datasets (18 and 16 channels for PD and ASD, respectively), we have modified the input layer of the GENet model to perform the training and classification process. For those two datasets, categorization performance decreases with the increase in the training batch size. For the PD dataset, the highest average

TABLE III
Five-Fold and LOOV Average Performance Results With Standard Deviation for the GENet Model on the Assessed Seven Datasets for Four Training Batch Sizes (Here, LAcc is the Accuracy for LOOV)

Batch Size		Datasets						
		SZ1	SZ2	MCI	ADHD	EP	PD	ASD
32	Sen%	98.49 ± 0.41	99.54 ± 0.14	99.98 ± 0.01	99.98 ± 0.03	100.00 ± 0.00	98.90 ± 0.54	98.91 ± 0.42
	Spec%	98.64 ± 0.65	99.45 ± 0.14	99.99 ± 0.01	99.97 ± 0.04	100.00 ± 0.00	99.77 ± 0.09	97.55 ± 0.65
	Prec%	99.10 ± 0.42	99.55 ± 0.11	99.99 ± 0.01	99.98 ± 0.03	100.00 ± 0.00	99.77 ± 0.08	98.96 ± 0.28
	F1	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.01	0.99 ± 0.00
	Acc%	98.55 ± 0.18	96.92 ± 12.22	**99.99 ± 0.01**	99.98 ± 0.01	**100.00 ± 0.00**	**99.34 ± 0.33**	**98.51 ± 0.43**
	LAcc%	**99.22 ± 2.06**	**99.50 ± 0.04**	**100.00 ± 0.00**	**100.00 ± 0.00**	99.51 ± 1.77	76.89 ± 21.10	90.88 ± 15.20
64	Sen%	98.71 ± 0.15	99.65 ± 0.11	99.98 ± 0.01	99.96 ± 0.05	100.00 ± 0.00	98.32 ± 0.92	98.73 ± 0.63
	Spec%	98.94 ± 0.24	99.50 ± 0.18	99.99 ± 0.01	99.95 ± 0.09	100.00 ± 0.00	99.20 ± 0.58	97.39 ± 1.24
	Prec%	99.30 ± 0.15	99.59 ± 0.14	99.99 ± 0.01	99.96 ± 0.07	100.00 ± 0.00	99.19 ± 0.59	98.90 ± 0.51
	F1	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.01	0.99 ± 0.01
	Acc%	98.80 ± 0.08	**99.58 ± 0.11**	**99.99 ± 0.01**	99.95 ± 0.06	**100.00 ± 0.00**	98.76 ± 0.72	98.34 ± 0.63
	LAcc%	99.16 ± 2.79	97.51 ± 11.91	**100.00 ± 0.00**	**100.00 ± 0.00**	99.51 ± 1.77	78.83 ± 20.62	91.12 ± 16.54
128	Sen%	99.10 ± 0.31	99.45 ± 0.14	100.00 ± 0.00	99.98 ± 0.03	100.00 ± 0.00	97.66 ± 0.60	98.77 ± 0.38
	Spec%	98.40 ± 0.29	99.54 ± 0.12	99.99 ± 0.01	100.00 ± 0.00	100.00 ± 0.00	99.05 ± 0.27	96.87 ± 2.03
	Prec%	98.95 ± 0.19	99.62 ± 0.10	99.98 ± 0.02	100.00 ± 0.00	100.00 ± 0.00	99.03 ± 0.20	98.67 ± 0.89
	F1	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.98 ± 0.00	0.99 ± 0.00
	Acc%	**98.82 ± 0.29**	99.49 ± 0.12	**99.99 ± 0.01**	**99.99 ± 0.02**	**100.00 ± 0.00**	98.37 ± 0.31	98.19 ± 0.66
	LAcc%	99.06 ± 2.58	96.96 ± 13.14	**100.00 ± 0.00**	**100.00 ± 0.00**	99.51 ± 1.77	**80.41 ± 19.14**	92.09 ± 12.89
256	Sen%	98.63 ± 0.42	99.53 ± 0.09	99.99 ± 0.02	99.94 ± 0.07	100.00 ± 0.00	97.34 ± 1.21	97.85 ± 0.98
	Spec%	98.35 ± 0.52	99.51 ± 0.13	99.98 ± 0.00	99.99 ± 0.03	99.95 ± 0.07	98.41 ± 1.12	96.41 ± 1.03
	Prec%	98.91 ± 0.34	99.59 ± 0.11	99.97 ± 0.01	99.99 ± 0.02	99.95 ± 0.07	98.39 ± 1.08	98.48 ± 0.44
	F1	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
	Acc%	98.51 ± 0.22	99.52 ± 0.03	99.98 ± 0.01	99.96 ± 0.03	99.97 ± 0.04	97.87 ± 0.76	97.42 ± 0.91
	LAcc%	98.66 ± 3.79	97.16 ± 12.34	**100.00 ± 0.00**	**100.00 ± 0.00**	99.51 ± 1.77	77.56 ± 19.97	**95.54 ± 9.71**

Note: The bold entries indicate the best result for that dataset.

accuracy 99.34($\pm$0.33)% is achieved using batch size 32 and for ASD it is 98.51($\pm$0.43)%. Highest sensitivity, specificity, precision, and F1 score values are 98.90($\pm$0.54)%, 99.77($\pm$0.09)%, 99.77($\pm$0.08)%, and 0.99($\pm$0.01), respectively, for the PD dataset and 98.91($\pm$0.42)%, 97.55($\pm$0.65)%, 98.96($\pm$0.28)%, and 0.99($\pm$0.00), respectively for ASD dataset. Fig. 7 shows a comparison graph of the evaluation parameters with SD for the proposed model.

In LOOV, we have achieved accuracy values close to the 5-fold CV accuracy for most of the datasets as given in Table III. For SZ1, LOOV accuracy is 98.82%, for SZ2, it is 99.50%, 100% for MCI, 100% for ADHD, 99.51% for EP, 95.54% for ASD, and 80.41% for PD. For SZ1, MCI, and ADHD, we have achieved better accuracy in LOOV than 5-fold CV. For SZ2 and EP, both the accuracies are closer, while for ASD and PD, the LOOV accuracy is lower than the 5-fold accuracy due to the smaller number of channels, but it is still comparatively good classification accuracy for an unknown test subject.

The ROC curve is a good indicator of the classifier's performance. If the curve of a classifier is close to the point (0, 1), then it is considered to be a good classifier, while if the curve is close to or below the diagonal line, then it is considered to be a poor classifier. We have plotted the ROC curve of the GENet model for the seven tested datasets, as shown in Fig. 8. From the figure, we can see that for EP, MCI, and ADHD datasets, the proposed model achieved perfect classification performance, and for other datasets, the performance is near perfect. Such

large areas under the ROC curves also prove the scalability of the proposed EEG classification framework.

2) Multiclass Classification: To further assess the performance of the proposed GENet model, we have performed a multiclass classification using four datasets among the seven tested datasets that have similar data sampling records. For this purpose, we have used the SZ 2, MCI, EP, and ADHD datasets and modified the final layer of the GENet model to perform a five-class (SZ versus MCI versus EP versus ADHD versus Normal) categorization task. This multiclass classification task is also evaluated using the five-fold cross-validation technique. Performance of this classification task is also measured using the five parameters as done for binary classification. Table IV summarizes the performance result of the GENet model for multiclass classification over five-fold and for four different batch sizes (32, 64, 128, and 256).

From Table IV, we can see that for multiclass categorization using GENet, increasing the batch size increases the performance of the model. The proposed model has achieved an accuracy of 99.75($\pm$0.07)% with batch size 32 which has increased to 99.81($\pm$0.04)%, 99.83($\pm$0.02)%, and 99.84($\pm$0.05)% for batch size 64, 128, and 256, respectively. To further inspect the impact of batch sizes on the individual folds, we have plotted the sensitivity, specificity, precision, and accuracy values as a spider plot in Fig. 9.

In Fig. 9, each axis of the graphs represents a testing fold, and the four colored areas represent four different tested batch sizes.
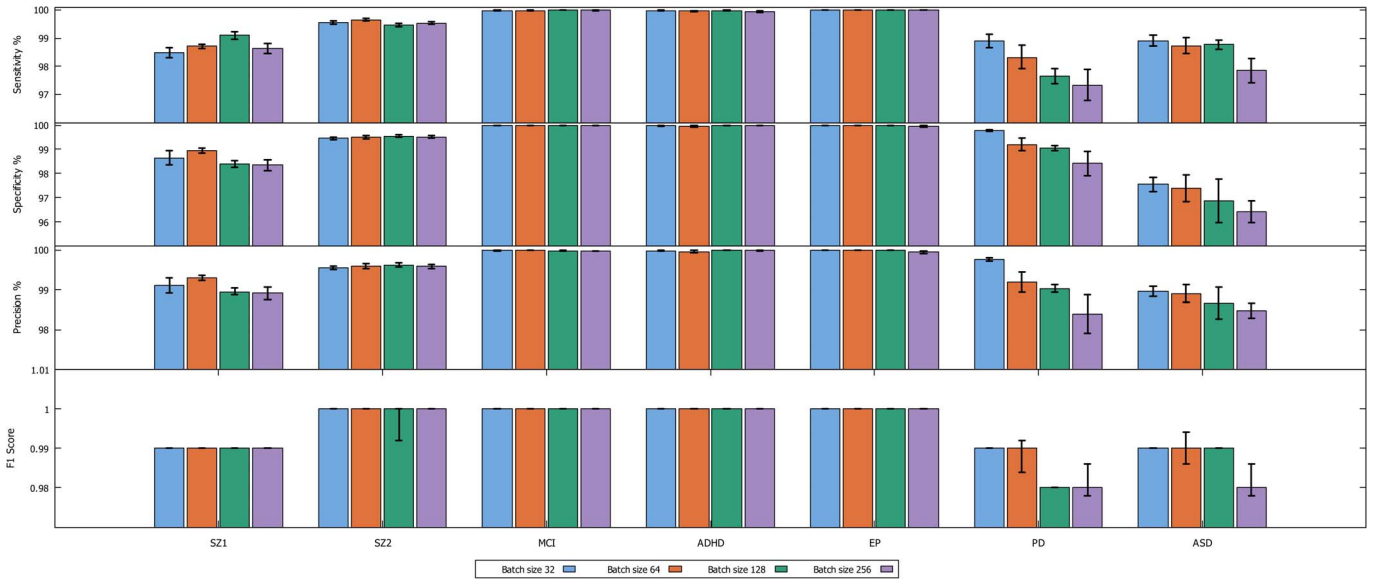
Fig. 7. Comparison graph of the four evaluation parameters (Sen, Spec, Prec, and F1) with standard deviation for the GENet model. Each subplot represents an evaluation parameter where *x*-axis has the seven assessed datasets. For each dataset, we have four bars for four different training batch sizes (32, 64, 128, and 256).
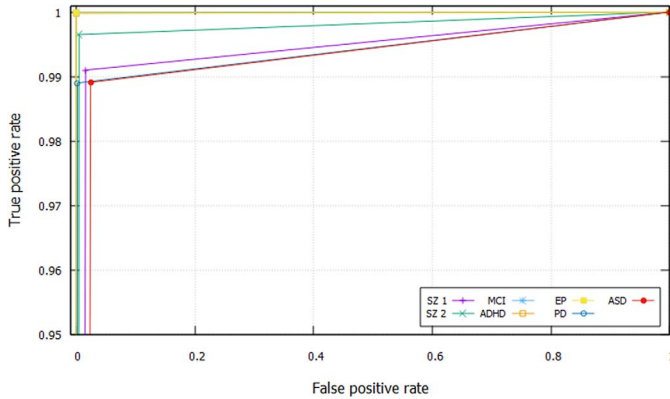

Fig. 8. ROC graph of the GENet model on tested datasets.

From Fig. 9, we can see that for a single fold, highest sensitivity of 99.87% is achieved for fold 5 of batch size 256 and fold 2 of batch size 64 while the lowest sensitivity value of 99.55% is obtained for fold 1 of batch size 32. In case of specificity, the highest and lowest values are 99.96% for fold 2 with batch size 256 and 99.88% for fold 1 with batch size 32, respectively. Fold 4 of batch size 64 has produced the highest precision of 99.92% and fold 5 of batch size 32 has given the lowest precision value of 99.64%. A single fold highest accuracy of 99.88% is achieved for fold 2 of batch size 256 and lowest of 99.68% is obtained for both the fold 1 and 5 of batch size 32.

D. Discussion

In this study, a generic framework using the GENet model is proposed to classify NDs using EEG data. GENet is a CNN-based model that is designed to take raw EEG signal data as input and train its internal layers with the significant features of

TABLE IV
FIVE-FOLD AVERAGE PERFORMANCE RESULTS OF THE GENet MODEL ON
MULTICLASS CLASSIFICATION

Batch Size		Sen%	Spec%	Prec%	F1	Acc%
256	Normal	99.86	99.83	99.84	1.00	
	SZ	99.59	99.93	99.60	1.00	
	MCI	99.95	99.99	99.96	1.00	**99.84**
	EP	99.76	99.99	99.90	1.00	
	ADHD	99.90	99.99	99.92	1.00	
	AVG	**99.81**	**99.94**	**99.84**	**1.00**	
128	Normal	99.86	99.80	99.81	1.00	
	SZ	99.46	99.94	99.68	1.00	
	MCI	99.96	99.98	99.94	1.00	**99.83**
	EP	99.79	99.99	99.89	1.00	
	ADHD	99.95	99.98	99.86	1.00	
	AVG	**99.80**	**99.94**	**99.84**	**1.00**	
64	Normal	99.86	99.78	99.80	1.00	
	SZ	99.43	99.93	99.62	1.00	
	MCI	99.94	99.99	99.96	1.00	**99.81**
	EP	99.79	99.99	99.89	1.00	
	ADHD	99.90	99.99	99.88	1.00	
	AVG	**99.78**	**99.94**	**99.83**	**1.00**	
32	Normal	99.78	99.74	99.76	1.00	
	SZ	99.27	99.91	99.51	1.00	
	MCI	99.91	99.98	99.94	1.00	**99.75**
	EP	99.78	99.99	99.87	1.00	
	ADHD	99.92	99.96	99.60	1.00	
	AVG	**99.73**	**99.92**	**99.74**	**1.00**	

Note: The bold entries indicate the average values of different evaluation parameters.

the data to perform classification tasks. To reduce the manual process in the classification steps, we have tried to develop a DL-based system so that it will extract and classify the features
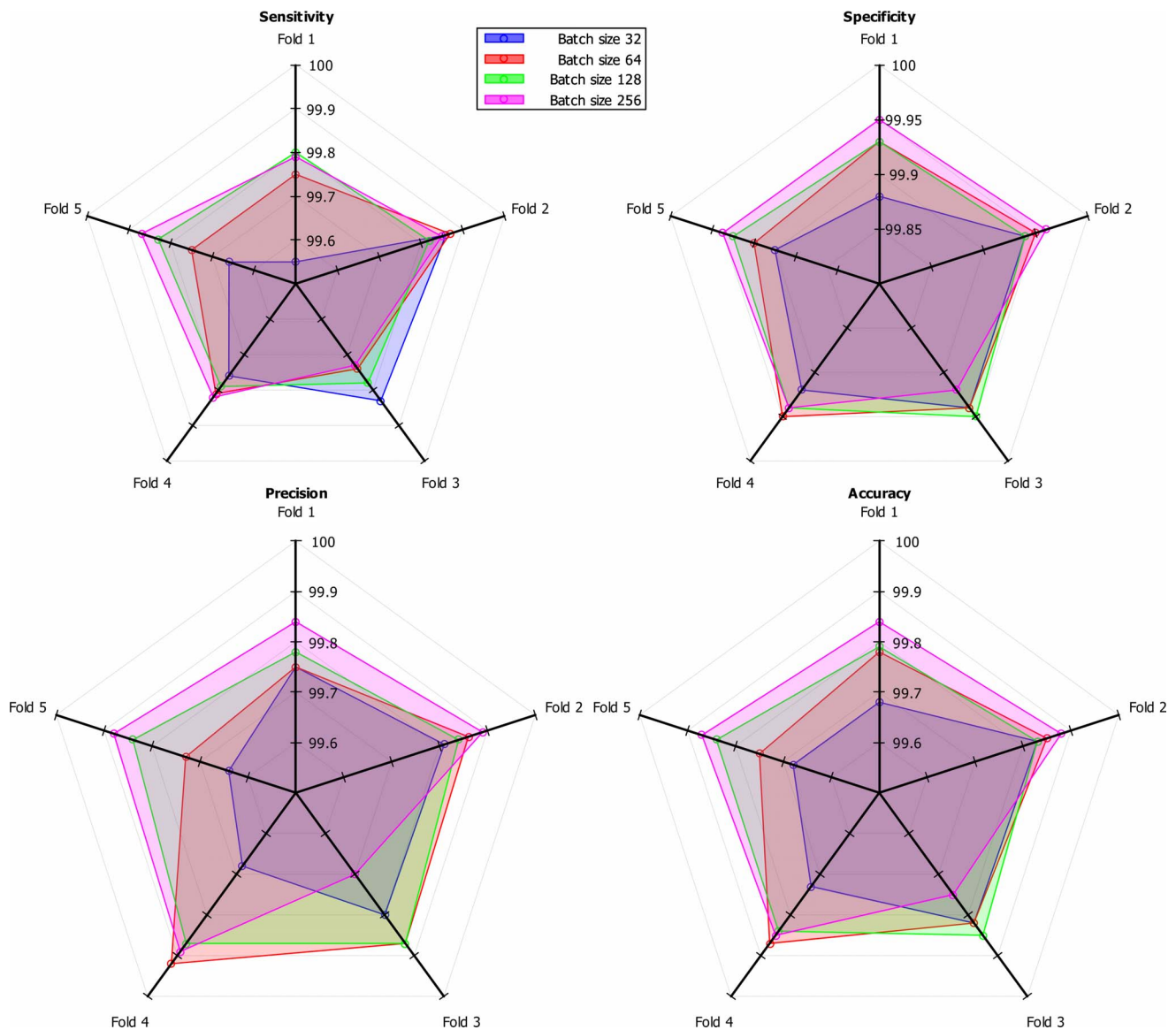
Fig. 9. Spider plot visualization for fold wise sensitivity, specificity, precision, and accuracy comparison of the GENet model for multiclass classification with different tested batch sizes. Data have five folds (Fold 1,2,..,5) and each polygon is a multivariate data point for a training batch size.

automatically. We have also tried to minimize the preprocessing steps by just segmenting the signals into small chunks and reducing the recording channels to make the EEG signals input ready for the GENet model. To prove the multi disease scalability of the proposed GENet model, we have evaluated it using seven different EEG datasets from six distinct NDs. We have tested two different datasets (SZ1 and SZ2) of the same disease (SZ) as well as five other datasets from five different diseases (MCI, ADHD, EP, PD, and ASD).

1) Ablation Study: Generally, an ablation study is a series of experiments in which parts of a ML system are removed or replaced to assess the effect of these parts on the system's performance. To validate the structure of the proposed GENet model, we have conducted ten different ablation studies using datasets SZ1 and SZ2, considering that the model will behave similarly for other datasets. In this ablation study, we have

used the results of the GENet model on datasets SZ1 and SZ2 for batch sizes 64 and 128, respectively, as the baseline result and performed those ablation studies using the same setup. To conduct the ablation process, we have divided the proposed model into four different blocks for ease of experimentation. As shown in Table II, every dropout layer divides the model into a block, so the first block consists of the first two Conv2D layers, the maxpooling2D layer, and the dropout layer. The second block consists of the second two Conv2D layers, the second maxpooling2D layer, the second dropout layer, and so on. The ablation experiments settings and the obtained results are given in Table V.

Different types of ablation processes were tested, including adding or removing the blocks, increasing or decreasing the filters of the convolution layers of different blocks, and adding convolution layers in different blocks. From Table V, we can

TABLE V
ABLATION STUDY RESULTS ON DATASET SZ1 AND SZ2

Ablation Methods	Accuracy %	
	SZ1	SZ2
Baseline (no ablation)	**98.82**	**99.58**
Removed 2nd block	97.57	99.13
Removed 3rd block	97.40	99.07
Removed both 2nd and 3rd block	81.05	97.63
Added a duplicate of 2nd block before 3rd block	98.12	98.94
Doubled the filters in first block	98.13	98.95
Doubled the filters in 2nd block	98.05	98.96
Doubled the filters in last block	98.16	98.97
Halved the filters in first block	98.13	99.03
Halved the filters in 2nd block	98.37	99.09
Halved the filters in last block	98.10	98.94
Added a conv2d layer in first block	98.14	98.99
Added a conv2d layer in second block	98.09	98.98
SGDM optimizer used in place of Adam	98.62	99.45
AlexNet	91.62	99.03
ResNet18	93.83	98.94

TABLE VI
TIME-COMPLEXITY ANALYSIS OF THE PROPOSED GENet MODEL FOR THE
TWO SZ DATASETS TESTED WITH DIFFERENT BATCH SIZE CONFIGURATIONS

Dataset	Batch Size	Time /Epoch	Training		Validation	
			Acc.%	Loss	Acc.%	Loss
SZ1	32	6s	98.46	0.05	98.55	0.05
	64	5s	98.94	0.03	98.80	0.05
	128	4s	98.96	0.03	98.82	0.05
	256	4s	98.77	0.04	98.52	0.06
SZ2	32	14s	99.49	0.01	99.50	0.04
	64	6s	99.56	0.01	99.58	0.02
	128	4s	99.57	0.01	99.49	0.02
	256	4s	99.63	0.01	99.52	0.02

TABLE VII
ACCURACY COMPARISON FOR DIFFERENT DATA
AUGMENTATION TECHNIQUES ON SZ1 AND
SZ2 DATASETS

Data Augmentation Methods	Accuracy %	
	SZ1	SZ2
No augmentation	99.58	98.82
Multiplied signal (Multi)	99.88	99.67
Adding noise (Noise)	99.76	99.37
Flipping the data (Flip)	99.66	99.27
Frequency shifting (Freq)	99.82	99.47
Noise + Flip	99.83	99.53
Noise + Multi	99.86	99.70
Flip + Freq	99.84	99.56

TABLE VIII
COMPARISON OF PROPOSED GENet MODEL WITH EXISTING
STATE-OF-THE-ART (SoA) STUDY USING SAME DATASETS

Dataset	SoA Authors	SoA Accuracy %	Our Accuracy %
SZ1	Xiaojun et al. [45]	92.00	**98.82**
SZ2	Mehmet et al. [46]	99.47	**99.58**
MCI	Siuly et al. [8]	98.78	**99.99**
ADHD	Ali et al. [47]	89.70	**99.99**
EP	Tawhid et al. [40]	98.79	**100.00**
PD	Anjum et al. [33]	85.70	**99.77**
ASD	Tawhid et al. [21]	99.15	**98.51**

Note: The bold entries indicate the accuracy of the proposed model.

see that adding or removing the blocks has a negative impact on the accuracy of the proposed system. For example, removing both the 2nd and 3rd blocks drops the accuracy from 98.82% to 81.05% for SZ1, while for SZ2 it drops from 99.58% to 97.63%. In the case of adding a single convolution layer, adding it in the first block gives a better result than adding it in the second block. On the other hand, changing the filters in the different blocks has a different effect. We have tried both halving and doubling the filters in each block. Among the three blocks, doubling the filters in the last block produces the best results, while halving the filters in the second block produces the best results. We have also tested the SGDM optimizer in place of Adam and got an accuracy of 98.62% and 99.45% for SZ1 and SZ2, respectively. Moreover, we have also tested and reported on two popular CNN models, AlexNet and ResNet18. All the tested ablation methods prove that the proposed GENet model gives a better result than the other tested models.

2) Time Complexity Analysis: Table VI shows the time–complexity analysis of the proposed GENet model on the two tested SZ datasets. From the table, we can see that, with the increase of the training batch size, the time per epoch decreases, but there is not much of a steady pattern in the accuracy and loss values of training and validation. Yet, considering the time complexity analysis, batch size 128 may gain good classification accuracy with a smaller training time.

3) Data Augmentation: Data augmentation is a quick way to add more labeled data to train a network and has been used extensively in the context of DL [40]. Due to the structure of the EEG signal, relatively few data augmentation methods can be applied while maintaining equal power, frequency, and spatial components. We have employed the same seven data augmentation methods as the authors in [40]: multiplication, frequency shift, adding noise, flipping data, and a combination of those four approaches on the SZ1 and SZ2 datasets, and the obtained results on those augmented data are given in Table VII.

Table VII shows that data augmentation improves the proposed model's performance and demonstrates its robustness to perturbations.

4) Comparison with Existing Studies: Finally, to compare the performance of the proposed GENet model with existing state-of-the-art (SoA) research work that have used the same datasets as we have used in this study, we have listed the SoA works with our accuracy in Table VIII. From Table VIII, we can see that, for datasets SZ1, SZ2, MCI, ADHD, EP, and PD, our proposed model has outperformed the SoA's accuracy. For the ASD dataset, our proposed model produces an accuracy close to the SoA, which is maybe due to the fact that it has fewer data channels (16) than other datasets (19) for which this GENet model is proposed.

IV. CONCLUSION

Here, we have proposed a DL-based generic framework for classifying ND from EEG data. First, we have preprocessed the EEG signal data and segmented the EEG signals into short time fragments. We have tested three different time segments (1, 2, and 3 s) to check the impact of the segment length on the detection process. Then, we have proposed a CNN model named GENet to classify the segmented signals into healthy or disordered groups. To assess the performance and scalability of the proposed framework, we tested it on seven different EEG datasets from six different NDs and performed extensive experimental work using 5-fold cross-validation.

Among the tested segment lengths, 1 s data segment gives the highest classification accuracy compared to the other two time segments. The proposed GENet model offers a higher classification performance for the seven tested datasets. For six of the seven tested datasets, our proposed model achieved higher accuracy than state-of-the-art work using those datasets. We have achieved an accuracy of 98.82%, 99.58%, 99.99%, 99.99%, 100%, 99.77%, and 98.51% for datasets SZ1, SZ2, MCI, ADHD, EP, PD, and ASD, respectively. We have also tested the proposed model for multiclass classification using four (SZ2, MCI, EP, and ADHD) of the seven datasets to perform a five-class (SZ versus MCI versus EP versus ADHD versus Normal) classification task and achieved an accuracy of 99.84%. Moreover, we have analyzed the t-SNE images of the proposed GENet model to check the extracted features' plotting and found that those features are clustered into separable classes. We have also completed several ablation studies to validate the proposed GENet model.

Finally, the findings show that this method is versatile and may be applied to multidisease classification tasks using EEG data and other signal-processing tasks. Despite this, the framework's high classification accuracy suggests that a fully automated, multifunctional, and effective neural rehabilitation system can be developed using the framework. As this study used a small dataset, a large number of EEG signals will be used to evaluate system performance in the future. We will also work on finding the important channels that contribute best to the classification process and changing the GENet model to work on those reduced channels.

REFERENCES

[1] World Health Organization, *Neurological Disorders: Public Health Challenges*. Switzerland: World Health Organization, 2006.

[2] World Health Organization et al., "Optimizing brain health across the life course: Who position paper," 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240054561

[3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Feb. 2008.

[4] J. Yin, J. Cao, S. Siuly, and H. Wang, "An integrated MCI detection framework based on spectral-temporal analysis," *Int. J. Automat. Comput.*, vol. 16, no. 6, pp. 786–799, 2019.

[5] S. Siuly, O. F. Alcin, V. Bajaj, A. Sengur, and Y. Zhang, "Exploring hermite transformation in brain signal analysis for the detection of epileptic seizure," *IET Sci., Meas. Technol.*, vol. 13, no. 1, pp. 35–41, 2018.

[6] Ö. F. Alcin, S. Siuly, V. Bajaj, Y. Guo, A. Sengu, and Y. Zhang, "Multi-category EEG signal classification developing time-frequency texture features based fisher vector encoding method," *Neurocomputing*, vol. 218, pp. 251–258, Dec. 2016.

[7] S. Siuly, S. K. Khare, V. Bajaj, H. Wang, and Y. Zhang, "A computerized method for automatic detection of schizophrenia using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2390–2400, Nov. 2020.

[8] S. Siuly et al., "A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 1966–1976, Sep. 2020.

[9] V. Geraedts et al., "Machine learning for automated EEG-based biomarkers of cognitive impairment during deep brain stimulation screening in patients with Parkinson's disease," *Clin. Neurophysiol.*, vol. 132, no. 5, pp. 1041–1048, 2021.

[10] S. K. Khare, V. Bajaj, and U. R. Acharya, "SPWVD-CNN for automated detection of schizophrenia patients using EEG signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[11] R. Jana and I. Mukherjee, "Deep learning based efficient epileptic seizure prediction with EEG channel optimization," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102767.

[12] S. K. Khare, V. Bajaj, and U. R. Acharya, "Pdcnnet: An automatic framework for the detection of Parkinson's disease using EEG signals," *IEEE Sensors J.*, vol. 21, no. 15, pp. 17017–17024, May 2021.

[13] B. Ari, N. Sobahi, Ö. F. Alcin, A. Sengur, and U. R. Acharya, "Accurate detection of autism using Douglas-Peucker algorithm, sparse coding based feature mapping and convolutional neural network techniques with EEG signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105311.

[14] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, Jan. 2019.

[15] H. Chen, Y. Song, and X. Li, "A deep learning framework for identifying children with adhd using an EEG-based brain network," *Neurocomputing*, vol. 356, pp. 83–96, Sep. 2019.

[16] S. Siuly, Y. Li, and Y. Zhang, "EEG signal analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 141–4, 2016.

[17] M. Tawhid, N. Ahad, S. Siuly, K. Wang, and H. Wang, "Brain data mining framework involving entropy topography and deep learning," in *Proc. Australas. Database Conf.,* Switzerland: Springer, 2022, pp. 161–168.

[18] M. N. A. Tawhid, S. Siuly, K. Wang, and H. Wang, "Automatic and efficient framework for identifying multiple neurological disorders from EEG signals," *IEEE Trans. Technol. Soc.*, vol. 4, no. 1, pp. 76–86, Mar. 2023.

[19] M. N. A. Tawhid, S. Siuly, and H. Wang, "Diagnosis of autism spectrum disorder from EEG using a time–frequency spectrogram image-based approach," *Electron. Lett.*, vol. 56, no. 25, pp. 1372–1375, 2020.

[20] A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin, and M. Elsharnouby, "Focus: Detecting adhd patients by an EEG-based serious game," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1512–1520, Jul. 2018.

[21] M. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG," *Plos one*, vol. 16, no. 6, 2021, Art. no. e0253094.

[22] W. Hussain, M. T. Sadiq, S. Siuly, and A. U. Rehman, "Epileptic seizure detection using 1 d-convolutional long short-term memory neural networks," *Appl. Acoust.*, vol. 177, 2021, Art. no. 107941.

[23] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2019.

[24] S. L. Oh et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 10927–10933, 2020.

[25] M. N. A. Tawhid, S. Siuly, E. Kabir, and Y. Li, "Exploring frequency band-based biomarkers of EEG signals for mild cognitive impairment detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 189–199, Dec. 2024.

[26] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGnet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[27] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[28] O. Mecarelli, "Electrode placement systems and montages," in *Proc. Clin. Electroencephalography*. Switzerland: Springer, 2019, pp. 35–52.

[29] B. Roach. "EEG data from basic sensory task in schizophrenia." Accessed: Feb. 2021. [Online]. Available: https://www.kaggle.com/broach/button-tone-sz

[30] E. Olejarczyk and W. Jernajczyk, "Graph-based analysis of brain connectivity in schizophrenia," *PLoS One*, vol. 12, no. 11, 2017, Art. no. e0188629.

[31] M. Kashefpoor, H. Rabbani, and M. Barekatain, "Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features," *J. Med. Signals Sensors*, vol. 6, no. 1, 2016, Art. no. 25.

[32] A. Pereira and J. Fiel, "Resting-state interictal EEG recordings of refractory epilepsy patients," Netherlands: Mendeley Data 2019. [Online]. Available: https://data.mendeley.com/datasets/6hx2smc7nw/1

[33] M. F. Anjum, S. Dasgupta, R. Mudumbai, A. Singh, J. F. Cavanagh, and N. S. Narayanan, "Linear predictive coding distinguishes spectral EEG features of parkinson's disease," *Parkinsonism Related Disorders*, vol. 79, pp. 79–85, Oct. 2020.

[34] M. J. Alhaddad et al., "Diagnosis autism by fisher linear discriminant analysis flda via EEG," *Int. J. Bio-Sci. Bio-Technol.*, vol. 4, no. 2, pp. 45–54, 2012.

[35] A. Motie Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi, "EEG data for adhd / control children," IEEE Dataport, 2020. [Online]. Available: https://dx.doi.org/10.21227/rzfh-zn36

[36] M. J. Rivera, M. A. Teruel, A. Maté, and J. Trujillo, "Diagnosis and prognosis of mental disorders by means of EEG and deep learning: A systematic mapping study," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 1–43, 2022.

[37] Z. Aslan and M. Akin, "Automatic detection of schizophrenia by applying deep learning over spectrogram images of EEG signals," *Traitement du Signal*, vol. 37, no. 2, pp. 235–244, 2020.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[39] M. Tawhid, N. Ahad, S. Siuly, K. Wang, and H. Wang, "Data mining based artificial intelligent technique for identifying abnormalities from brain signal data," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, Switzerland: Springer, 2021, pp. 198–206.

[40] M. N. A. Tawhid, S. Siuly, and T. Li, "A convolutional long short-term memory based neural network for epilepsy detection from EEG," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[41] M. N. A. Tawhid, S. Siuly, K. Wang, and H. Wang, "Textural feature based intelligent approach for neurological abnormality detection from brain signal data," *Plos One*, vol. 17, no. 11, 2022, Art. no. e0277555.

[42] S. Siuly, X. Yin, S. Hadjiloucas, and Y. Zhang, "Classification of thz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers," *Comput. Methods Programs Biomed.*, vol. 127, pp. 64–82, Apr. 2016.

[43] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 945–954, Dec. 2020.

[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[45] X. Yu, M. Z. Aziz, M. T. Sadiq, K. Jia, Z. Fan, and G. Xiao, "Computerized multidomain EEG classification system: A new paradigm," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 3626–3637, Aug. 2022.

[46] M. Baygin, O. Yaman, T. Tuncer, S. Dogan, P. D. Barua, and U. R. Acharya, "Automated accurate schizophrenia detection system using collatz pattern technique with EEG signals," *Biomed. Signal Process. Control*, vol. 70, 2021, Art. no. 102936.

[47] A. Ekhlasi, A. M. Nasrabadi, and M. Mohammadi, "Classification of the children with adhd and healthy children based on the directed phase transfer entropy of EEG signals," *Frontiers Biomed. Technol.*, vol. 8, no. 2, pp. 115–122, 2021.