SPWVD-CNN for Automated Detection of Schizophrenia Patients Using EEG Signals

Smith K. Khare, *Graduate Student Member, IEEE*, Varun Bajaj, *Senior Member, IEEE*, and U. Rajendra Acharya, *Senior Member, IEEE*

***Abstract*— Schizophrenia (SZ) is a psychiatric disorder characterized by cognitive dysfunctions, hallucinations, and delusions, which may lead to lifetime disability. Detection and diagnosis of SZ by visual inspection is subjective, difficult, and time-consuming. Electroencephalogram (EEG) signals are widely used to detect the SZ as they reflect the conditions of the brain. Conventional machine learning methods involve many lengthy manual steps, such as decomposition, feature extraction, feature selection, and classification. In this article, automated identification of SZ is proposed using a combination of time–frequency analysis and convolutional neural network (CNN) to overcome the limitations of feature extraction-based methods. Three press button tasks are analyzed to segregate normal subjects and SZ patients. The EEG signals are subjected to continuous wavelet transform, short-time Fourier transform, and smoothed pseudo-Wigner–Ville distribution (SPWVD) techniques to obtain scalogram, spectrogram, and SPWVD-based time–frequency representation (TFR) plots, respectively. These 2-D plots are fed to pretrained AlexNet, VGG16, ResNet50, and CNN. We have obtained an accuracy of 93.36% using the SPWVD-based TFR and CNN model. In comparison to the benchmark AlexNet, ResNet50, and VGG16 networks, the developed CNN model with four convolutional layers not only requires fewer learnable parameters but also is computationally efficient and fast. This clearly indicates that our proposed method combining the SPWVD-CNN model has performed better than the state-of-the-art transfer learning techniques. Our developed model is ready to be tested with more EEG data and can aid psychiatrists in their diagnosis.**

***Index Terms*— Convolutional neural networks (CNNs), electroencephalogram (EEG) signals, scalogram, schizophrenia (SZ) detection, spectrogram, time–frequency representation (TFR).**

I. Introduction

SCHIZOPHRENIA is a common neurological disorder, which occurs mostly during adulthood. Approximately 28 million people, i.e., 1% of the global population, are affected by schizophrenia (SZ). According to the American psychiatry association, patients experiencing SZ have fantasies (hallucinations), disorganized speech, delusions, and so forth [1]. Prognosis of SZ requires a long-term economic burden and medication [2]. Electrocardiogram and electromyogram signals have been used for the identification of SZ [3], [4]. However, such models did not yield high performance. Lack of standardized tools and methods has reduced the performance of early identification of SZ. Electroencephalogram (EEG) signals provided significant information over other techniques for the diagnosis of SZ [5]. The EEG signals are noninvasive, economical, and nonradioactive. Hence, they have been widely used to detect brain abnormalities these days [6]–[8].

To date, many methods have been proposed by researchers for the classification of SZ using EEG signals. Power and low-recurrence frequency content of alpha-band in [9] and gamma-band by filtering in [10] have been employed to detect SZ. Giannitrapani and Kayton [11] and Morstyn *et al.* [12] used spectral analysis and fast Fourier transform (FFT) techniques to detect SZ. Microstate k-means clustering technique has been utilized to detect SZ in [13]. In [14]–[16], numerous features, namely band power, fractal dimension, and autoregressive (AR) model coefficients, coupled with linear discriminant analysis (LDA), multi-LDA, support vector machine (SVM), and adaptive boosting classifiers, have been used to detect SZ automatically. Independent component analysis (ICA) and time–frequency analysis using the Stockwell transform have been utilized in [17], to discover the progressions of rhythms during SZ. In [18], recurrence plots and nonlinear classification techniques have been used for feature extraction and classifying SZ patients automatically. An 11-layered deep convolutional neural network (CNN) model has been used in the automatic detection of SZ [19]. An optimal wavelet-based feature extracted from the subbands has been classified with k-nearest neighbor (kNN) classifier [20].

In [21] and [22], multiple rhythms have been extracted by filtering and mean coherence is classified using the SVM classifier to automatically detect SZ. Features extracted by Higuchi, Katz, and Petrosian methods in [23] have been classified using boosted LDA. Bandpass filter and SVM with leave-one-out cross validation is used to detect SZ automatically in [24]. Multiset canonical correlation analysis with recursive feature elimination and SVM has been used for the discrimination of SZ in [25]. In [26], the detection of SZ has been accomplished using the Higuchi fractal dimension, Kolmogorov complexity, and entropy-based features with SVM. The short-time Fourier transform (STFT) with

Manuscript received February 24, 2021; accepted March 25, 2021. Date of publication April 2, 2021; date of current version April 16, 2021. The Associate Editor coordinating the review process was Dr. Lorenzo Ciani. *(Corresponding author: Varun Bajaj.)*

Smith K. Khare and Varun Bajaj are with the Electronics and Communication Discipline, Indian Institute of Information Technology Design and Manufacturing, Jabalpur 482005, India (e-mail: smith7khare@gmail.com; varunb@iiitdmj.ac.in).

U. Rajendra Acharya is with the Division of Electronic and Computer Engineering, School of Engineering, Ngee Ann Polytechnic, Singapore 599489, also with the Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore 599494, and also with the Department of Biomedical Informatics and Medical Engineering, Asia University, Taichung 41354, Taiwan.

Digital Object Identifier 10.1109/TIM.2021.3070608

a sliding window technique has been used in [27], to distinguish the SZ patients automatically. The Riemannian geometric and weighted minimum distance means have been used to detect SZ automatically in [28]. The Hilbert–Huang transform, ICA, and local discriminant bases in [29] and FFT, eigenvector, wavelet transform, and AR methods in [30] have been used for the recognition of SZ patients. Features extracted by an inherent spatial pattern of network have been classified using SVM and LDA classifiers in [31]. In [32], energy and power of EEG signals have been used for the identification of SZ. Random forest (RF) and SVM have been used for the classification of ϵ-complexity vector functions in [33]. Empirical wavelet transform (EWT) and empirical mode decomposition have been used for feature extraction [34], [35]. Several machine learning algorithms have been used for classifying SZ patients and normal subjects.

The methods proposed in this literature involve the identification of SZ by separation of rhythms, extraction of features (time, frequency, entropy, spectral, and nonlinear), and classification. Selection of suitable features is tedious and time-consuming. Also, traditional approaches require multiple tuning parameters for decomposition, feature extraction, and classification. However, the empirical selection of these tuning parameters may cause information loss and increased misclassification [36]–[38]. Hence, there is a need to develop a novel method to detect SZ accurately at an early stage. To address this, temporal details of nonstationary EEG signals in terms of time–frequency–energy are obtained by time–frequency representation (TFR) techniques. The images obtained by smoothed pseudo-Wigner–Ville distribution (SPWVD), STFT, and continuous wavelet transform (CWT) are classified using CNNs. STFT and CWT with CNN have been employed in multiple studies to detect abnormalities, such as drowsiness, motor imagery task, and seizures [39]–[41]. Recently, the SPWVD-CNN combination has been used for the automated classification of emotions with EEG signals [42]. In this work, we are using spectrogram, scalogram, and SPWVD in combination with CNNs to detect SZ automatically using EEG signals. To the best of our knowledge, we are the first group to use this methodology for the automated detection of SZ using EEG signals.

In this article, the spectrograms, scalograms, and SPWVD-based TFR using CNNs are proposed for SZ detection. The TFRs are obtained by STFT, CWT, and SPWVD methods, respectively. The TF plots are given as input to different CNN models. Features are extracted and classified by three pretrained CNNs and CNN.

The major contributions of the proposed work are summarized as follows.

1) SZ affects an individual's motor and auditory ability. Hence, analysis of press button tasks involving three different conditions is studied and identified the best condition to be used for the automated detection of SZ.
2) Time, frequency, and nonlinear domain features have been used to analyze the EEG data. In this work, we have used TFR plots obtained from EEG signals using STFT, CWT, and SPWVD.
3) Identification, evaluation, selection, and classification of suitable features is tedious and time-consuming. Hence, a novel deep learning (DL) technique using CNN is developed using TFR plots as input signals.

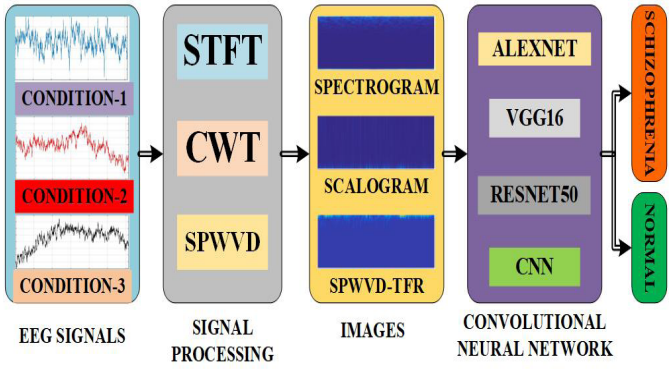

Fig. 1. Proposed methodology.

The organization of the remaining article is as follows. The methodology is described in Section II. Sections III and IV presents the results and a discussion of the proposed method. Finally, the gist of this article is presented in Section V.

II. Methodology

The methodology consists of obtaining EEG data set, transformation techniques, and CNNs. The EEG signals of three conditions are used in the data set. Three transformation techniques, namely STFT, CWT, and SPWVD, are employed to obtain the spectrograms, scalograms, and SPWVD-based TFRs. The different time–frequency plots are fed to the various types of CNNs for automated detection of SZ. The proposed methodology is shown in Fig. 1.

A. Data Set

The EEG data obtained from Kaggle contains basic sensory tasks of 81 human subjects. The details of a data set can be found in [43] and [44]. The data set contains 49 SZ patients and 32 healthy control (HC) subjects. There are 67 male and 14 female subjects with 39 years of average age. Three simple press button conditions have been employed: 1) immediately generate a tone after pressing a button; 2) listening to the same tone passively; and 3) pressed a button without the generation of a tone. The EEG data have been digitized at 1024 Hz. The signals were preprocessed with a digital bandpass filter to remove different artifacts with a frequency range of 0.5 and 15 Hz [45]. The EEG signals were recorded from 64 scalp and 8 external sites. A total of 3108, 3015, and 3111 EEG signals belonging to conditions 1–3 for normal subjects are obtained. For SZ patients, a total of 4608, 4506, and 4520 signals that belong to conditions 1–3, respectively, are created. Each signal consists of 3072 samples. The typical EEG signal of conditions 1–3 for normal and SZ patients is shown in Fig. 2. It is evident from this figure that normal EEG signals in condition 1 and SZ patients provide significant variation in amplitudes, whereas the EEG signals of conditions 2 and 3 have low amplitude range.

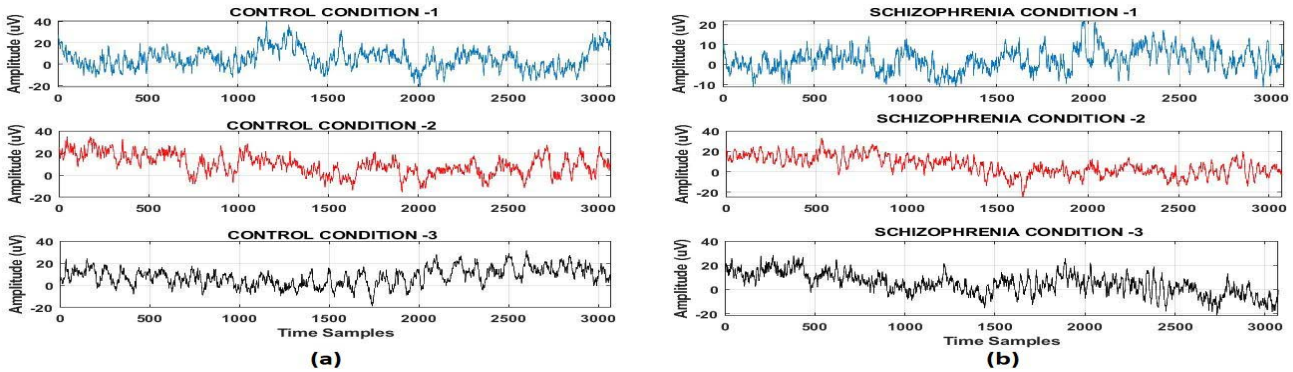

Fig. 2. Typical EEG signals at various conditions: (a) normal and (b) SZ.

B. Transformation Techniques

It is difficult to analyze the EEG signals as it contains profoundly oscillatory amplitudes and fluctuating frequency components. To get a thorough insight into the signals, it needs to be translated from one domain to another. Transformation techniques enable capturing the details of time-domain signals into frequency domain simultaneously. TFR captures the spectral variation of the signal over time. This article employed STFT, CWT, and SPWVD to convert the signal into its TF representation.

1) Short-Time Fourier Transform: It is an advanced version of Fourier transform, which evaluates the change in frequency of signal over time. This is achieved by cutting the time-domain signal into finite blocks, and then, the Fourier transform is evaluated on each block. It is due to the property of dividing a signal into blocks, STFT has gained numerous applications using EEG signals [27], [39], [40]. Fourier transform is applied to each block of time-domain signal using the window function to extract various properties. STFT is symmetric bandpass filtering, which is evenly spaced in frequency. The STFT of signal $z(t)$ is given by [46]

$$Z_w(t, f) = \int_0^T z(\tau) w(t-\tau) \exp^{-j2\pi f \tau} \mathrm{d}\tau. \tag{1}$$

The spectrogram is defined as a squared magnitude of STFT. The spectrogram is mathematically expressed as $S_w(t, f) = |Z_w(t, f)|^2$, where $w(t)$ is the window function. The length of the windows for each block must be equal. A signal $z(t)$ is assumed nearly stationary in the duration of window.

2) Continuous Wavelet Transform: Wavelet is an oscillating bandpass filter having finite energy and zero average function. The wavelet transform can be obtained with CWT or discrete wavelet transform. The mathematical representation for CWT of signal $z(t)$ is denoted by [47]

$$W_{\alpha,\beta}[z(t)] = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} z(t) \psi^* \left(\frac{t-\beta}{\alpha} \right) \tag{2}$$

where α is a scaling parameter that controls the spread of function and β denotes the time shift and $\psi((t-\beta)/\alpha)$ is the shifted and scaled version of mother wavelet $\psi(t)$. The advantage of CWT lies in its ability to alter the size of the window. Analysis of low-frequency components can be obtained by a large window and vice versa.

3) Smoothed SPWVD: It is another powerful transformation technique that provides information on time–frequency localization of signal energy. It enables the evaluation of higher harmonics and time-domain localization. It uses time and frequency windows to reduce the cross-term interference. The expression of SPWVD can be written as [48]

$$\mathrm{SPWVD}(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v\left(\frac{\tau}{2}\right) v^*\left(-\frac{\tau}{2}\right) u(t-t') \ldots \times z\left(t' + \frac{\tau}{2}\right) z^*\left(t' - \frac{\tau}{2}\right) \exp^{-j2\pi f t} \mathrm{d}t' \mathrm{d}\tau \tag{3}$$

where the cross-term reducing window in time and frequency domains is $v(t)$ and $u(t)$. The length of time and frequency windows can be set independently to adjust the temporal and frequency resolution. STFT assumes the signal to be stationary over a duration of window. The CWT produces a cross-term interference in frequency domain. The tradeoff between time–frequency localization and time–frequency cross term produced by spectrogram and scalogram can be overcome using SPWVD [48]. The typical TFR plots obtained from STFT, CWT, and SPWVD are shown in Fig. 3.

C. Convolutional Neural Network

Traditional feature extraction and classification methods require qualitative and quantitative analysis for decision-making. The CNN is a DL technique that has become popular due to its ability of automatic deep feature extraction and classification. The CNN uses convolutional operation instead of matrix multiplication with input as an image and identifies various aspects to differentiate one class from another. The CNN is a multilayer network that comprised of interconnected neurons. These neurons are trained to extract and classify the features. The basic architecture of CNN is composed of input, hidden, and output layers. The hidden layer is a combination of multiple convolutional layers, pooling layers, and fully connected layers [49]. The convolutional and pooling layers are responsible for deep feature extraction, and a fully connected layer is used for classification.

1) *Convolutional Layer:* It is the core block of CNN, which consists of a set of kernels known as learnable parameters. The feature maps from input or previous layers are convolved with learnable kernels to extract

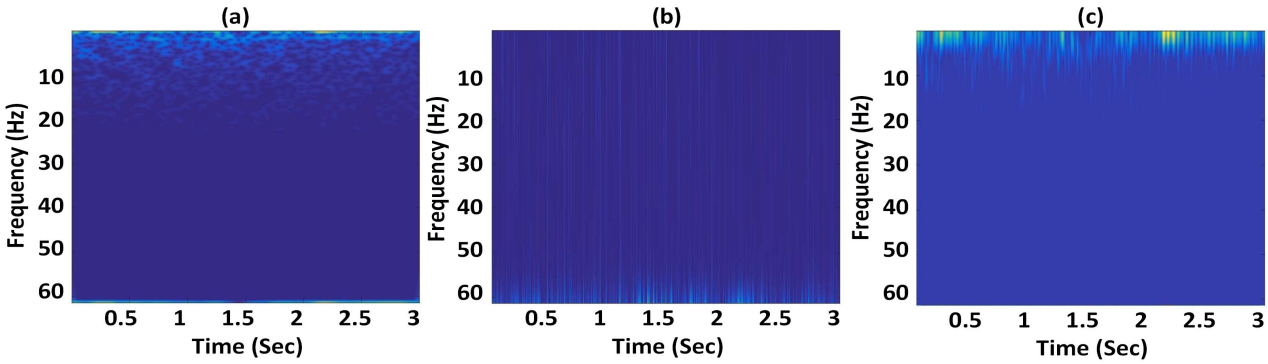

Fig. 3. Sample TFR plots obtained from (a) STFT, (b) CWT, and (c) SPWVD.

the depth of input volume. The convolutional operation is expressed as

$$X * Y(m,n) = \int X(i,j)Y(m+i,n+j). \quad (4)$$

Stride is used to move the kernels with a fixed number of pixels. Usually, to maintain the size of image, zero paddings are also applied. Nonlinearity in the network is achieved by the activation function in combination with convolution. In this work, rectified linear unit (ReLu) is used as an activation function.

2) *Pooling Layer:* The pooling or subsampling layer is mainly used to reduce the dimension of output maps. Downsampling operation is used to accomplish the dimensionality reduction. The number of input and output maps remains the same. For any N input maps, the reduced output maps are denoted as

$$y_l^k = f\left(a_l^k \text{down}\left(y_l^{k-1}\right) + b_l^k\right) \quad (5)$$

where down(.) is the pooling function, and multiplicative and additive bias terms are denoted by a_l^k and b_l^k, respectively. Mostly, two types of pooling are performed, namely average and max pooling.

3) *Fully Connected Layer:* The output of pooling layer is feed as an input to a fully connected layer. The scores of each class from the extracted features from preceding convolutional layers are computed by this layer. It converts 2-D data into 1-D. The final layer is a vector of scalar values that are used to classify the data using the softmax classification layer.

The above layers can be combined to build a CNN. One can add or drop the number of convolutional, pooling, and fully connected layers to meet the desired performance. Many pre-trained models have been developed recently. It transfers the previously learned knowledge from one domain to another for feature extraction and classification. In this article, AlexNet, VGG16, and ResNet50 networks are employed for feature extraction and classification. The details of these networks are given in [49]. Training and testing of CNN require tuning of a number of hyperparameters. Classification accuracy can be controlled by numerous tuning parameters. Hence, a CNN with a fewer number of convolutional, pooling, and a fully connected layer is designed. The proposed CNN consists of four convolutional layers, two pooling layers, a dropout layer, and two fully connected layers [42]. The architecture of the proposed CNN is shown in Fig. 4.

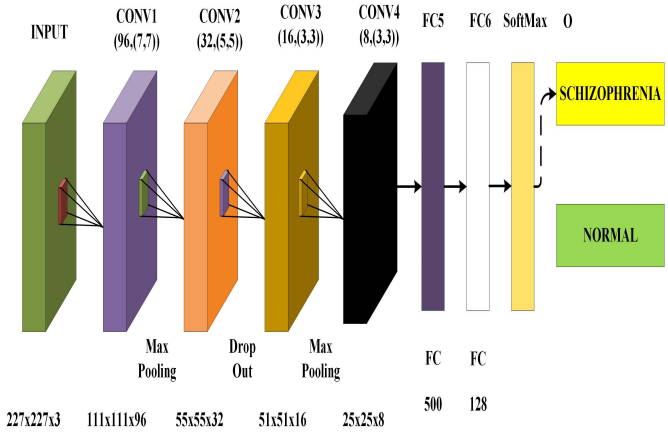

Fig. 4. Developed CNN model.

III. Results

Conventional methods require signal analysis, feature extraction, selection, and classification steps to develop the automated SZ detection system. Multiple parameters are to be tuned to obtain the desired performance. Conventional approaches are time-consuming and burdensome. Many times optimum performance is not achieved even after multiple trials. To overcome this, a CNN-based method is proposed for automatic feature extraction and classification. Three conditions of 32 normal and 49 SZ patients are considered in this study. The EEG signals are transformed to the T-F domain using STFT, CWT, and SPWVD methods. The same parameters are used for each method to maintain the uniformity. The STFT is used with a nonoverlapping Hamming window with a sampling frequency of 256 Hz. The Morlet wavelet is used to obtain the scalograms using CWT [40]. Kaiser window is selected in frequency and time domains to reduce the cross term to obtain TFR-based on SPWVD. The medium length of window = 31 is chosen as a smaller window result in poor resolution, while a larger window increases the size of image.

The spectrograms, scalograms, and SPWVD-based TFRs are fed as an input to AlexNet, VGG16, ResNet50, and CNN.

In this work, we have used the same hyperparameters of AlexNet, VGG16, ResNet50, and CNN. The training and testing is carried out using tenfold cross validation. Nine parts have been used for training and the rest is utilized for validation, and this is repeated ten times. The learning rate of each weight is scaled by the Adam optimizer. A number of epochs are set to 7 and the batch size is selected to be 50. A validation frequency is chosen as 3 and a learning rate of 10^{-04}. AlexNet and CNN take the input size of 227×227, whereas ResNet50 and VGG16 take input size as 224×224. Hence, all images are resized to 227×227 and 224×224.

The classification accuracies obtained for conditions 1–3 using AlexNet, VGG-16, ResNet50, and CNN are shown in Table I. The best performance is obtained for condition 1 using TFR-based on SPWVD. The highest accuracies are 93.33%, 93.09%, 93.34%, and 93.36% for AlexNet, VGG16, ResNet50, and CNN, respectively. The TFR-based on SPWVD provided more representative information compared with spectrogram and scalogram due to its ability to reduce the cross-term independently in time and frequency simultaneously. The developed CNN model with four convolutional layers is able to extract representative information from TFR based on SPWVD. The best classification of HC and SZ is obtained for condition 1 followed by conditions 2 and 3. It can be concluded that the likelihood of detecting SZ increases when subjects perform both auditory and motor tasks concurrently due to differences in SZ EEG patterns relative to HC subjects. On the other hand, EEG patterns of SZ and HC subjects may remain unchanged or may show very little improvement when performing auditory or motor tasks separately, due to which the accuracy for classification of conditions 2 and 3 is much lower. Table I shows that the developed CNN has achieved better classification accuracy compared with other deep networks such as AlexNet, VGG16, and ResNet50 even after fine-tuning the hyperparameters such as batch size and learning rate.

The training and testing accuracies obtained using TFR for each iteration with AlexNet, VGG16, ResNet, and CNN for condition 1 are shown in Fig. 5(a)–(d), respectively. It can be interpreted from the figure that AlexNet requires few iterations to reach its maximum performance but provided a fluctuating performance, especially during final iterations of each epoch. ResNet50 and CNN provided stable performance with the least variation after the model is settled to its maximum performance after few hundred iterations. Finally, the VGG-based model fluctuated during middle iterations of training and validation process making this model unstable.

To understand the proposed method more clearly, six performance parameters are obtained for condition 1. Six performance parameters, namely sensitivity (SEN), accuracy (ACC), specificity (SPE), F-1 score, Mathew's correlation coefficient (MCC), and precision (PRC), are computed to evaluate the proposed method. The mathematical formulation of performance parameters are given in the following:

TABLE I

CLASSIFICATION ACCURACY (%) OBTAINED USING DIFFERENT CONDITIONS, METHODS, AND CNN-BASED MODELS

Method	Spectrogram	Scalogram	SPWVD
Condition 1			
AlexNet	78.40	90.28	93.33
VGG16	78.56	89.20	93.09
ResNet50	79.18	90.19	93.34
CNN	79.17	90.64	93.36
Condition 2			
AlexNet	50.53	57.67	61.31
VGG16	51.03	58.29	60.53
ResNet50	52.53	59.40	61.19
CNN	51.97	58.82	61.76
Condition 3			
AlexNet	52.13	53.24	65.16
VGG16	52.41	52.39	65.67
ResNet50	51.88	54.85	66.52
CNN	51.49	54.63	66.64

TABLE II

PERFORMANCE PARAMETERS FOR DIFFERENT CNN MODELS USING SPWVD-BASED TFR

Parameters	AlexNet	ResNet50	VGG16	CNN
ACC (%)	93.33	93.34	93.09	93.36
SEN (%)	94.89	94.60	93.94	94.25
SPE (%)	91.07	91.49	91.82	92.03
MCC (%)	86.17	86.17	85.62	86.19
F1-Score	0.944	0.944	0.942	0.945
PRC (%)	93.88	94.23	94.53	94.66

$$
\begin{aligned}
\text{Recall} = \text{SEN} &= \frac{T_p}{T_p + F_n} \\
\text{ACC} &= \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \\
\text{SPE} &= \frac{T_n}{T_n + F_p} \\
\text{MCC} &= \frac{T_p * T_n - F_p * F_n}{\sqrt{(T_p + F_p)(T_n + F_n)(T_n + F_p)(T_p + F_n)}} \\
\text{PRC} &= \frac{T_p}{T_p + F_p} \\
F\text{-1} \quad \text{Score} &= \frac{2 * \text{PRC} * \text{Recall}}{\text{PRC} + \text{Recall}} \qquad (6)
\end{aligned}
$$

where T_p, F_n, T_n, and F_p are the true positive, false negative, true negative, and false positives of classification, respectively.

Table II shows the performance parameters for condition 1 using AlexNet, ResNet50, VGG16, and CNN models with SPWVD-based TFR. The proposed CNN has provided higher ACC, SPE, MCC, F1-Score, and PRC over AlexNet, VGG16, and ResNet50, while SEN of SPWVD-CNN is higher than that of VGG16 and lesser than that of AlexNet and ResNet50. The summary of various tuning parameters used in different CNN-based models is shown in Table III.

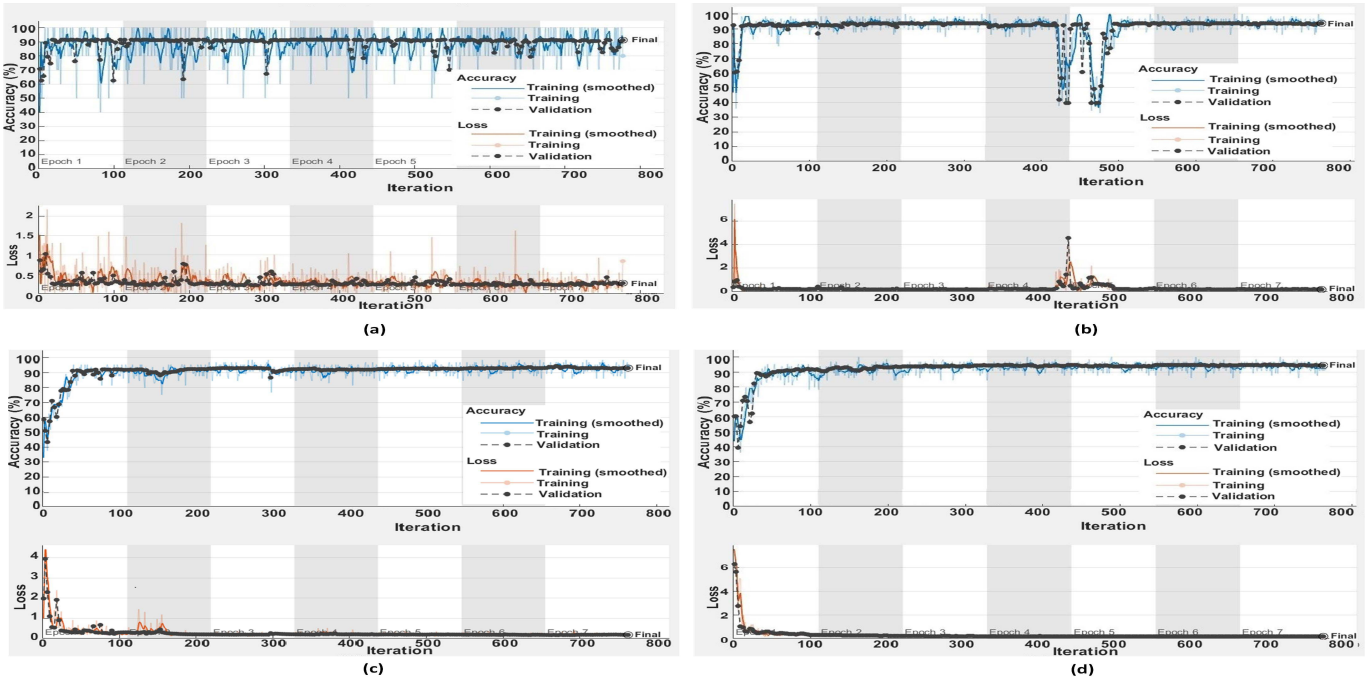

Fig. 5. Training and testing accuracies obtained for various epochs with TFR using different models: (a) AlexNet, (b) VGG16, (c) ResNet50, and (d) CNN.

TABLE III
SUMMARY OF TUNING PARAMETERS USED FOR CNN-BASED MODELS

Parameters	CNN	AlexNet	ResNet50	VGG16
Convolutional layers	4	5	50	16
Dense layer	2	3	1	2
Size of filters	7,5,3	11,5,3	7,3,1	3
Stride	1,2	1,4	1,2	1
Approx. LP	1M	61M	25.5M	138M
Approx. TRFEF	32min	37min	164min	400min
Approx. TT	300min	360min	1600min	3970min

Approx. - approximate, LP - learning parameters, TRFEF - time required for each fold, TT - total time

TABLE IV
CONFUSION MATRIX OBTAINED USING CNN

Class	SZ	Normal	Total
SZ	4362 (94.66%)	246 (5.34%)	4608
Normal	266 (8.56%)	2842 (91.44%)	3108
Total	4628	3088	7716

It is evident from the table that the developed CNN model has used fewer learnable parameters compared with AlexNet, ResNet50, and VGG16 models. This justifies that our developed CNN model is computationally efficient, fast, and less complex. The results of Tables II and III show that the proposed four-layered CNN architecture is a suitable choice for an automated SZ detection.

The confusion matrix obtained for SZ patients and normal control classes for condition 1 using CNN is shown in Table IV.

IV. DISCUSSION

The proposed method is compared with state-of-the-art techniques. The summary of comparison is shown in Table V. Sabeti *et al.* [14] used autoregression, band power, and wavelet energy coefficients in combination with LDA. The method used plus-L-minus R and bidirectional search technique for channel selection and genetic algorithm for the selection of suitable features. Their developed system achieved an accuracy of 84.62% and 88.23% using bidirectional and plus-L-minus-R feature selection techniques. The single-stage feature selection classification accuracy is not as good as two stages, but it requires higher computational time (because of larger feature vector) and also unable to obtain the optimal values of L and R. The Higuchi, Katz, and Petrosian feature extraction methods have been used by Boostani *et al.* [23]. Their features with the LDA classifier obtained an accuracy of 87.5%. However, these methods in [14] and [23] captured EEG signals from the subjects without performing any auditory or motory tasks but only studied certain portion of scalp with a fixed number of electrodes. Johannesen *et al.* [21] used a filtering method to separate rhythms from the EEG signals of SZ and normal controls who completed a Sternberg working memory task. The rhythms with SVM achieved an accuracy of 87% using frontal alpha and theta at baseline during retrieval as primary classifiers of diagnosis. Shim *et al.* [24] used stimuli for the auditory oddball paradigm, and the participants are required to press a response button when target tones are presented. The significant features extracted from different frequency bands using bandpass filtering with the SVM classifier yielded 88.24% accuracy.

TABLE V
SUMMARY OF COMPARISON WITH STATE-OF-THE-ART TECHNIQUES

Author	EEG dataset	Feature extraction	Classifier	ACC (%)
Sabeti et al. [14]	20 HC subjects and 20 SZ patients	Autoregression, band power, and wavelet energy	LDA	88.23
Boostani et al. [23]	18 HC subjects and 13 SZ patients	Higuchi, Katz, and Petrosian methods	LDA	87.5
Johannesen et al. [21]	12 HC subjects and 40 SZ patients	Rhythms separated using filtering	SVM	87
Shim et al. [24]	34 HC subjects and 34 SZ patients	Bandpass filtering	SVM	88.24
Sui et al. [25]	53 HC subjects and 48 SZ patients	Multiset canonical correlation analysis	SVM	74
Piryatinska et al. [33]	39 HC subjects and 45 SZ patients	ϵ-complexity vector	RF	83.6
Jahmunah et al. [18]	14 HC subjects and 14 SZ patients	Nonlinear feature extraction	SVM	92.91
Oh et al. [19]	14 HC subjects and 14 SZ patients	CNN	CNN	98.07
Sharma et al. [20]	14 HC subjects and 14 SZ patients	optimal wavelet-based l1-norm	kNN	99.21
Zhang [50]	32 HC subjects and 49 SZ patients	Event related potential	RF	80.16 80.68 81.1
Khare et al. [35]	32 HC subjects and 49 SZ patients	Empirical wavelet transform	SVM	88.7
Siuly et al. [34]	32 HC subjects and 49 SZ patients	Empirical mode decomposition	EBT	89.59
Proposed	**32 HC subjects and 49 SZ patients**	**Smoothed pseudo-Wigner Ville distribution**	**AlexNet** **ResNet50** **VGG16** **CNN**	**93.33** **93.34** **93.09** **93.36**

Sui *et al.* [25] used resting states EEG recording and a multiset canonical correlation analysis to extract the features. These features coupled with the SVM classifier obtained 74% accuracy. Piryatinska *et al.* [33] used complexity vector functions with RF classifier obtained an accuracy of 83.6%. Jahmunah *et al.* [18] used nonlinear feature extraction and classification techniques. Their method yielded an accuracy of 92.91% with 12 features and SVM classifier. Oh *et al.* [19] developed 11-layered CNN model for subjective and non-subjective screening. Their network provided an accuracy of 98.07% and 81.26% or nonsubject and subject-based testing. Recently, Sharma and Acharya [20] have detected SZ using optimal wavelet-based l1-norm features extracted from single-channel EEG with an accuracy of more than 99%. The aforementioned methods have used different data sets with fewer subjects and focused on resting-state potential for the automatic detection of SZ.

Zhang [50] used different feature combinations to classify SZ using event-related potential (ERP) from nine channels. The subtracted signals of condition 1 from condition 3 are compared with condition 2 for both P200 (positive peak at 200-ms post-ERP), N100 (negative peak at 100-ms post-ERP), and peak-to-peak P300-N100. A combination of two, five, and eight features have been classified with an RF classifier and yielded an accuracy of 80.16%, 80.68%, and 81.1%, respectively. Khare *et al.* [35] used EWT and extracted various features (kurtosis, variance, root mean square, mean, and minima). These features with the SVM classifier obtained an accuracy of 88.7%. Siuly *et al.* [34] used the EMD-based decomposition for feature extraction. The features (maxima, minima, standard deviation, Hjorth activity, and first quantile) with ensemble bagged tree (EBT) classifier yielded an accuracy of 89.59%. Our proposed method used STFT, CWT, and SPWVD methods to obtain spectrogram, scalogram, and SPWVD-based TFR plots. The SPWVD-based TFR plots yielded an accuracy of 93.33%, 93.34%, 93.09%, and 93.36% using AlexNet, ResNet50, VGG16, and CNN models, respectively. It is clear from Table V that SPWVD-CNN has yielded the highest accuracy compared with other existing methods using the same data set. The existing methods used smaller data sets that are localized to study the resting state potential signals from the subjects. Also, traditional machine learning methods are used for signal decomposition, feature identification, feature evaluation, feature selection, and classification. This requires qualitative and quantitative analysis and hence may be time-consuming and exhaustive. Therefore, this study explored EEG signals obtained during auditory and motor tasks to automatically detect SZ using TFR and CNN. As subjects perform auditory and motor tasks simultaneously, the likelihood of SZ detection increases in contrast to performing auditory and motor tasks separately. The TFR obtained from SPWVD helped to obtain more representative information compared with STFT and CWT due to reduction in cross term of time–frequency, yielding in higher classification performance with CNN. The CNN makes it simpler to automate the extraction and classification of features, thereby significantly reducing manual effort. Moreover, our developed CNN model is simple, fast, and yielded better performance using fewer learnable parameters. Thus, we are able to obtain good performance using less complex CNN model, and hence, the computational time is less. The only issue is due to the use of convolutional operations and deep features, and memory requirements of the system may increase (compared with conventional feature extraction and classification techniques) [51].

The merits of our proposed method are as follows.

1) The highest detection accuracy of 93.36% is achieved with CNN architecture of only four convolutional layers.
2) The generated model is robust as it is developed using tenfold cross validation.

3) System is simple, less complex, fast, and completely automated as there is no need to perform separate feature extraction and classification.

The limitations of our developed model are given in the following.

1) Employed empirical selection of parameters.
2) Memory requirement has increased.

V. Conclusion

This article investigates the performance of various time–frequency methods and CNN-based models. Three conditions have been tested to detect SZ patients from normal subjects. Generating the tone after pressing the button provided better classification of SZ and HC compared with other two conditions. We have obtained better accuracy using SPWVD than STFT and CWT methods. This indicates that the TFR obtained by SPWVD provides in-depth information than spectrogram and scalogram plots. The proposed CNN model with four convolutional layers not only achieved the highest accuracy among other CNN-based networks but also is computationally efficient and fast. In future, we intend to use our developed model to automatically discriminate five different types of SZ accurately, which can aid the clinicians to provide the proper treatment immediately.

References

[1] S. B. Guze, "Diagnostic and statistical manual of mental disorders, 4th ed. (DSM-IV)," *Amer. J. Psychiatry*, vol. 152, no. 8, p. 1228, 1995.

[2] H. A. Whiteford, A. J. Ferrari, L. Degenhardt, V. Feigin, and T. Vos, "The global burden of mental, neurological and substance use disorders: An analysis from the global burden of disease study 2010," *PLoS ONE*, vol. 10, no. 2, pp. 1–14, Feb. 2015.

[3] M. T. Blom *et al.*, "Brugada syndrome ECG is highly prevalent in schizophrenia," *Circulat., Arrhythmia Electrophysiol.*, vol. 7, no. 3, pp. 384–391, Jun. 2014.

[4] L. Flyckt *et al.*, "Muscle biopsy, macro EMG, and clinical characteristics in patients with schizophrenia," *Biol. Psychiatry*, vol. 47, no. 11, pp. 991–999, Jun. 2000.

[5] T. Miyauchi, K. Tanaka, H. Hagimoto, T. Miura, H. Kishimoto, and M. Matsushita, "Computerized EEG in schizophrenic patients," *Biol. Psychiatry*, vol. 28, no. 6, pp. 488–494, Sep. 1990.

[6] S. K. Khare, V. Bajaj, and G. R. Sinha, "Adaptive tunable Q wavelet transform-based emotion identification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9609–9617, Dec. 2020.

[7] S. Taran and V. Bajaj, "Sleep apnea detection using artificial bee colony optimize Hermite basis functions for EEG signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 608–616, Feb. 2020.

[8] L. Angrisani, P. Arpaia, D. Casinelli, and N. Moccaldi, "A single-channel SSVEP-based instrument with off-the-shelf components for trainingless brain-computer interfaces," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3616–3625, Oct. 2019.

[9] S. R. Sponheim, B. A. Clementz, W. G. Iacono, and M. Beiser, "Clinical and biological concomitants of resting state EEG power abnormalities in schizophrenia," *Biol. Psychiatry*, vol. 48, no. 11, pp. 1088–1097, Dec. 2000.

[10] B. J. Roach and D. H. Mathalon, "Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia," *Schizophrenia Bull.*, vol. 34, no. 5, pp. 907–926, Jul. 2008.

[11] D. Giannitrapani and L. Kayton, "Schizophrenia and EEG spectral analysis," *Electroencephalogr. Clin. Neurophysiol.*, vol. 36, pp. 377–386, Jan. 1974.

[12] R. Morstyn, F. H. Duffy, and R. W. McCarley, "Altered topography of EEG spectral content in schizophrenia," *Electroencephalogr. Clin. Neurophysiol.*, vol. 56, no. 4, pp. 263–271, Oct. 1983.

[13] D. Lehmann *et al.*, "EEG microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: A multi-center study," *Psychiatry Res., Neuroimaging*, vol. 138, no. 2, pp. 141–156, Feb. 2005.

[14] M. Sabeti, R. Boostani, S. D. Katebi, and G. W. Price, "Selection of relevant features for EEG signal classification of schizophrenic patients," *Biomed. Signal Process. Control*, vol. 2, no. 2, pp. 122–134, Apr. 2007.

[15] M. Sabeti, S. D. Katebi, R. Boostani, and G. W. Price, "A new approach for EEG signal classification of schizophrenic and control participants," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2063–2071, Mar. 2011.

[16] M. Sabeti, R. Boostani, and T. Zoughi, "Using genetic programming to select the informative EEG-based features to distinguish schizophrenic patients," *Neural Netw. World*, vol. 22, no. 1, pp. 3–20, 2012.

[17] Z. Dvey-Aharon, N. Fogelson, A. Peled, and N. Intrator, "Schizophrenia detection and classification by advanced analysis of EEG recordings using a single electrode approach," *PLoS ONE*, vol. 10, no. 4, Apr. 2015, Art. no. e0123033.

[18] V. Jahmunah *et al.*, "Automated detection of schizophrenia using nonlinear signal processing methods," *Artif. Intell. Med.*, vol. 100, Sep. 2019, Art. no. 101698. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365719303628

[19] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," *Appl. Sci.*, vol. 9, no. 14, p. 2870, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/14/2870

[20] M. Sharma and U. R. Acharya, "Automated detection of schizophrenia using optimal wavelet-based ι_1 norm features extracted from single-channel EEG," *Cognit. Neurodynamics*, pp. 1–4, Dec. 2020.

[21] J. K. Johannesen, J. Bi, R. Jiang, J. G. Kenney, and C.-M.-A. Chen, "Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults," *Neuropsychiatric Electrophysiol.*, vol. 2, no. 1, pp. 1–21, Dec. 2016.

[22] S. Aydin, N. Arica, E. Ergul, and O. Tan, "Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements," *Int. J. Neural Syst.*, vol. 25, no. 3, May 2015, Art. no. 1550010.

[23] R. Boostani, K. Sadatnezhad, and M. Sabeti, "An efficient classifier to diagnose of schizophrenia based on the EEG signals," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6492–6499, Apr. 2009.

[24] M. Shim, H.-J. Hwang, D.-W. Kim, S.-H. Lee, and C.-H. Im, "Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features," *Schizophrenia Res.*, vol. 176, nos. 2–3, pp. 314–319, Oct. 2016.

[25] J. Sui *et al.*, "Combination of FMRI-SMRI-EEG data improves discrimination of schizophrenia patients by ensemble feature selection," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 3889–3892.

[26] B. Thilakvathi, K. Bhanu, and M. Malaippan, "EEG signal complexity analysis for schizophrenia during rest and mental activity," *Biomed. Res.*, vol. 28, pp. 1–9, Jan. 2017.

[27] A. Bachiller *et al.*, "Decreased entropy modulation of EEG response to novelty and relevance in schizophrenia during a P300 task," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 265, no. 6, pp. 525–535, Sep. 2015.

[28] F. AliMardani, R. Boostani, and B. Blankertz, "Presenting a spatial-geometric EEG feature to classify BMD and schizophrenic patients," *Int. J. Adv. Telecommun., Electrotech., Signals Syst.*, vol. 5, no. 2, pp. 79–85, Mar. 2016.

[29] W. A. W. Azlan and Y. F. Low, "Feature extraction of electroencephalogram (EEG) signal—A review," in *Proc. IEEE Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2014, pp. 801–806.

[30] A. S. Al-Fahoum and A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *Int. Scholarly Res. Notices*, vol. 2014, pp. 1–7, 2014.

[31] F. Li *et al.*, "Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 594–602, Apr. 2019.

[32] S. Zhang, Q. Shini, and W. Wang, "Classification of schizophrenia's EEG based on high order pattern discovery," in *Proc. IEEE 5th Int. Conf. Bio-Inspired Comput., Theories Appl. (BIC-TA)*, Sep. 2010, pp. 1147–1149.

[33] A. Piryatinska, B. Darkhovsky, and A. Kaplan, "Binary classification of multichannel-EEG records based on the ϵ-complexity of continuous vector functions," *Comput. Methods Programs Biomed.*, vol. 152, pp. 131–139, Dec. 2017.

[34] 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.

[35] S. K. Khare, V. Bajaj, S. Siuly, and G. R. Sinha, "Classification of schizophrenia patients through empirical wavelet transformation using electroencephalogram signals," in *Modelling and Analysis of Active Biopotential Signals in Healthcare*, vol. 1. Bristol, U.K.: IOP Publishing, 2020, pp. 1–26, doi: 10.1088/978-0-7503-3279-8ch1.
[36] A. Lay-Ekuakille *et al.*, "Entropy index in quantitative EEG measurement for diagnosis accuracy," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1440–1450, Jun. 2014.
[37] S. K. Khare and V. Bajaj, "A facile and flexible motor imagery classification using electroencephalogram signals," *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, Art. no. 105722.
[38] S. K. Khare and V. Bajaj, "An evolutionary optimized variational mode decomposition for emotion recognition," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2035–2042, Jan. 2021.
[39] U. Budak, V. Bajaj, Y. Akbulut, O. Atila, and A. Sengur, "An effective hybrid model for EEG-based drowsiness detection," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7624–7631, Sep. 2019.
[40] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional neural network based approach towards motor imagery tasks EEG signals classification," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4494–4500, Jun. 2019.
[41] P. Z. Yan, F. Wang, N. Kwok, B. B. Allen, S. Keros, and Z. Grinspan, "Automated spectrographic seizure detection using convolutional neural networks," *Seizure*, vol. 71, pp. 124–131, Oct. 2019.
[42] S. K. Khare and V. Bajaj, "Time-frequency representation and convolutional neural network-based emotion recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 31, 2020, doi: 10.1109/TNNLS.2020.3008938.
[43] J. M. Ford, V. A. Palzes, B. J. Roach, and D. H. Mathalon, "Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone," *Schizophrenia Bull.*, vol. 40, no. 4, pp. 804–812, Jul. 2014.
[44] *Kaggle: Your Machine Learning and Data Science Community*. Accessed: Feb. 22, 2020. [Online]. Available: https://www.kaggle.com/broach/button-tone-sz
[45] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in EEG data," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1770–1779, Jul. 2017.
[46] S. Nawab, T. Quatieri, and J. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 4, pp. 986–998, Aug. 1983.
[47] A. Grossmann, R. Kronland-Martinet, and J. Morlet, "Reading and understanding continuous wavelet transforms," in *Wavelets*, J.-M. Combes, A. Grossmann, and P. Tchamitchian, Eds. Berlin, Germany: Springer, 1990, pp. 2–20.
[48] M. Szmajda, K. Gorecki, and J. Mroczka, "Gabor transform, Gabor-wigner transform and SPWVD as a time-frequency analysis of power quality," in *Proc. 14th Int. Conf. Harmon. Qual. Power (ICHQP)*, Sep. 2010, pp. 1–8.
[49] M. Z. Alom *et al.*, "The history began from AlexNet: A comprehensive survey on deep learning approaches," Mar. 2018, *arXiv:1803.01164*. [Online]. Available: https://arxiv.org/abs/1803.01164
[50] L. Zhang, "EEG signals classification using machine learning for the identification and diagnosis of schizophrenia," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4521–4524.
[51] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek, "Comparing different deep learning architectures for classification of chest radiographs," *Sci. Rep.*, vol. 10, no. 1, pp. 611–629, Dec. 2020.

Smith K. Khare (Graduate Student Member, IEEE) is currently a Research Scholar at the Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India. He has authored or coauthored 14 publications in various high impact factor, peer-reviewed journals, such as IEEE Transactions and Elsevier. He has published two papers in international conference. His research interests include biomedical signal processing, pattern recognition, brain–computer interface, machine learning, and deep neural networks.

Mr. Khare is serving as a Reviewer in IEEE, Elsevier, and several other reputed journals.

Varun Bajaj (Senior Member, IEEE) received the Ph.D. degree from the Indian Institute of Technology Indore, Indore, India, in 2014.

He has been working as a faculty at the Discipline of Electronics and Communication Engineering, Indian Institute of Information Technology, Design and Manufacturing (IIITDM), Jabalpur, India, since 2014. He has authored more than 110 research papers in various reputed international journals/conferences, such as IEEE Transactions, Elsevier, Springer, and IOP. He has edited books in IOP and CRC Publisher. The citation impact of his publications is around 2485 citations, H-index of 26, and i10 index of 60 (Google Scholar March 2021). He has guided six (03 completed and 3 Inprocess) Ph.D. scholars and five M.Tech. scholars. His research interests include biomedical signal processing, image processing, time–frequency analysis, and computer-aided medical diagnosis.

Dr. Bajaj was a recipient of various reputed national and international awards. He is an Associate Editor of IEEE SENSORS JOURNAL and the Subject Editor-in-Chief of *IET Electronics Letters*.

U. Rajendra Acharya (Senior Member, IEEE) received the M.Tech. degree from Mangalore University, Mangalore, India, in 1998, the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India, in 2002, the D.Eng. degree from Chiba University, Chiba, Japan, in 2008, and the D.Sc. degree from the AGH University of Science and Technology, Kraków, Poland, in 2020.

He is currently a Senior Faculty Member with Ngee Ann Polytechnic, Singapore. He is also an Adjunct Professor at the University of Malaya, Kuala Lumpur, Malaysia, and Asia University, Taichung, Taiwan, an Associate Faculty Member at the Singapore University of Social Sciences, Singapore, and an Adjunct Professor at the University of Southern Queensland, Toowoomba, QLD, Australia. He has published more than 500 papers, in refereed international SCI-IF journals (345), international conference proceedings (42), and books (17) with more than 38 500 citations in Google Scholar (with H-index of 101). He has worked on various funded projects, with grants worth more than 5.5 million SGD. He is ranked in the top 1% of the highly cited researchers for the last five consecutive years (2016–2020) in computer science according to the Essential Science Indicators of Thomson.

Dr. Acharya is on the editorial board of many journals and has served as a guest editor for many journals.