Sleep Apnea Detection From Variational Mode Decomposed EEG Signal Using a Hybrid CNN-BiLSTM

TANVIR MAHMUD[1], **ISHTIAQUE AHMED KHAN**[1], **TALHA IBN MAHMUD**[1],
SHAIKH ANOWARUL FATTAH[1], (Senior Member, IEEE),
WEI-PING ZHU[2], (Senior Member, IEEE), AND
M. OMAIR AHMAD[2], (Life Fellow, IEEE)

[1]Department of Electrical and Electronics Engineering (EEE), Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh
[2]Department of Electronics and Communication Engineering (ECE), Concordia University, Montréal, QC H3G 1M8, Canada

Corresponding author: Shaikh Anowarul Fattah (fattah@eee.buet.ac.bd)

ABSTRACT Sleep apnea, a severe sleep disorder, is a clinically complicated disease that requires timely diagnosis for proper treatment. In this paper, an automated deep learning-based approach is proposed for the detection of sleep apnea frames from electroencephalogram (EEG) signals. Unlike conventional methods of direct feature extraction from EEG signals, the variational mode decomposition (VMD) algorithm is utilized in the proposed method to decompose the EEG signals into a number of modes. Use of such decomposed EEG signals for feature extraction offers efficient processing of the variations introduced in the frequency spectrum during apnea events irrespective of particular patients. Afterward, a fully convolutional neural network (FCNN) is proposed to separately extract the temporal features from each VMD mode in parallel while maintaining their temporal dependencies. The FCNN block utilizes causal dilated convolutions with increasing dilation rates along with multiple kernel operations in convolutions. Subsequently, for further exploration of the inter-modal temporal variations, these extracted features from different EEG-modes are jointly optimized with a stack of bi-directional long short term memory (LSTM) layers. Hence, the trained and optimized network is capable of generating predictions of apnea frames during the evaluation phase. Contrary to other studies, this study is carried out in a subject independent manner where separate subjects are considered for training and testing. Additionally, a semi-supervised approach is explored where for facilitating better classification performance on a subject's frames, a small portion of the patient's data is included in training to leverage insight regarding the possible environmental variations. Extensive experimentations on three publicly available datasets provide average accuracy of 93.22%, 93.25% and 89.41% in the subject-independent cross-validation scheme.

INDEX TERMS Sleep apnea, variational mode decomposition, computer-aided diagnosis, neural network, EEG signal.

I. INTRODUCTION

Sleep apnea is one of the most prevalent and severe sleep disorders which causes restriction of airflow with repetitive interruptions of breathing during sleep. According to some studies, about $5 - 21\%$ of the general adult population are diagnosed with sleep apnea [1], [2]. The apnea events can lead to substantial harmful physiological disorders which have the potential to develop longstanding sequelae, for instance, hypertension and cardio-respiratory disorders including heart attacks [3], [4], impairment of neuropsychological competence [5], [6], depression [7] and early all-cause mortality [8]. An increased level of cyclic alternating pattern rate was observed among apnea patients in [9] leading to possible impairments in certain cognitive domains. In traditional polysomnography (PSG), an expert collects numerous physiological signals including electroencephalogram (EEG) from the patient during sleep, and apneic episodes in the readings

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

are later annotated manually. Shortage of sleep centers and sleep technologists [10] along with the inter-technologist scoring variability and other man-made errors [11], [12], are some of the major obstacles in proper diagnosis of apnea. An automated apnea detection scheme can provide a powerful tool for circumventing these human errors and infrastructural shortcomings.

Towards achieving this goal, various physiological signals, such as oxygen saturation, the variation of heart rate and respiratory effort along with varieties of bio-signals like ECG, EMG, and EOG have been used in numerous studies [13]–[15]. Due to the technical complications and high expenditures of acquiring multiple bio-signals along with the rapid development of wearable-wireless EEG acquisition systems [16], EEG based analysis has received special attention from researchers in recent times in numerous sleep-related problems [17]. In [18], the authors proposed Hermite coefficient based decomposition of EEG signal for apnea detection using artificial bee colony optimization. De-trended fluctuation analysis based study of non-linear behavior and power-law correlations of EEG signal is considered for apnea detection in [19].

According to [20]–[24], the spectral contents in various frequency bands of EEG signal are significantly different during apneic episodes compared to other non-apnea periods of sleep. Hence, a number of studies have been carried out to extract effective features for apnea detection from decomposed EEG data. Depending on the vigilance, frequency spectrum of the EEG signal is traditionally divided into five frequency bands, namely delta (0.25-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz) and beta (16-40 Hz). Since in these frequency bands there exists variations in frequency (Hz), amplitude and activity level, in various applications, such frequency divisions facilitate the feature extraction process [21], [25], [26]. Delta, theta and alpha bands correspond to deep sleep, mild sleep and relax state, respectively, while sigma and beta bands refer to alert states. In [26], it is shown that during apnea, the energy contents in various frequency bands change significantly with respect to non-apnea events, especially, at that time higher-frequency bands exhibit greater relative energy contribution than that of lower-frequency bands. In [27], increased delta and beta activity is reported for the obstructive sleep apnea patients with high Apnea Hypopnea Index (AHI) compared to the simple snoring group having lower AHI. In [28], relative decrease in delta band power is found during the apnea events. Instead of frequency division, empirical mode decomposition is also used to analyze the characteristics of the EEG signal during apnea events. In [22], it is shown that the use of variational mode decomposition with adaptive center frequency surpasses the conventional spectral division of EEG data with manual band-pass filtering due to its better capability in capturing the variation of EEG characteristics induced due to apneic events. All these approaches depend on complicated statistical hand-crafted feature extraction process, such as subframe based entropy and log-variance

extraction [21], [22], transition of inter-band energy contribution extraction [20], and different spectral band ratio monitoring [20]. Nevertheless, it becomes extremely difficult with these manual, hand-crafted schemes to incorporate the most optimum features from various complex temporal and inter-modal relationships of EEG data.

To the best of our knowledge, all other studies have been carried out in a subject-dependent manner where information of each subject is used for both training and testing which severely limits the applicability of these schemes in real-life test scenarios with unknown subjects. In this paper, a subject-independent automated approach is proposed for detection of apnea frames where a customized deep convolutional-BiLSTM neural network is used for efficient feature extraction from the EEG signal. Rather than using the raw EEG signal, the pre-processed signal is decomposed into a number of modes using variational mode decomposition (VMD) algorithm for adaptive variations in center-mode frequencies. Instead of extracting hand-crafted features from the VMD-decomposed EEG data, a deep fully convolutional network (FCNN) is proposed to independently extract the complex temporal features from each EEG-mode separately. During this operation their causal relationships are maintained in large-time windows using causal dilated convolutions. These FCNN modules project the temporal information of each EEG-mode into a smaller subspace that contains more generalized variations of features relating apnea episodes utilizing multiple kernel convolutions for efficient pooling. Thereafter, a stack of bi-directional long-short term memory (BiLSTM) layers is introduced to process the outputs of all FCNNs together for exploring the inter-modal temporal feature variations in different VMD modes. The whole process is schematically shown in Fig. 1. This approach automatically extracts the complicated temporal and inter-modal feature relationships of the EEG data that are mostly supposed to be affected during apnea events. Intensive experimentations carried out in a subject-independent manner provide outstanding performance that proves the robustness of the proposed scheme for real-world applications.

II. PROPOSED METHODOLOGY
The proposed scheme consists of a number of operations for precise recognition of apnea events. All of these are discussed in detail in the following subsections.

A. FRAME CREATION AND PROBLEM FORMULATION
The raw sampled and digitized EEG signals collected from patients are divided into a number of frames of uniform length to facilitate the processing with deep neural network. Frames are created with a predefined length from the continuous EEG data and its corresponding label is assigned according to the annotated apnea label at the middle of the respective frame. Subsequently, a predefined shift is considered in between subsequent frames to introduce overlapping samples in-between frames. Such overlapping increases the number
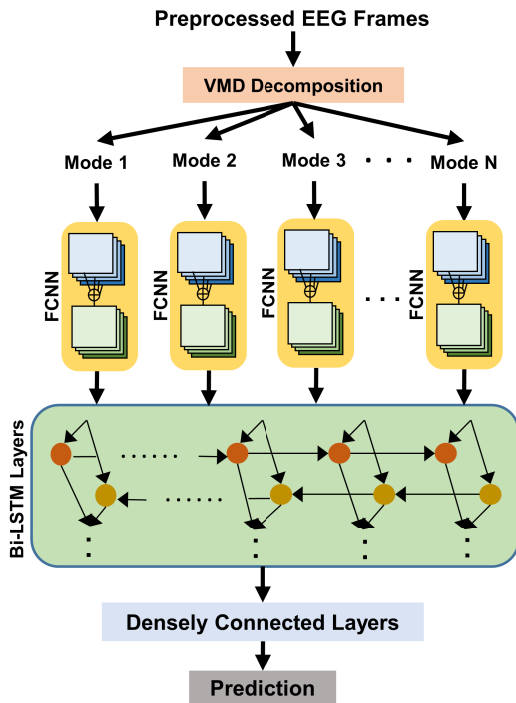
FIGURE 1. Schematic representation of the proposed workflow.

of available frames that can be used for training. Moreover, larger frame length can be considered for exploring long term dependencies of temporal features without a significant reduction in the number of training frames by increasing overlapping ratio. However, shorter frame length can also be used with a smaller overlapping ratio between subsequent frames.

Let's consider the set of extracted EEG frames as, $D = \{(\mathbf{x}_i, y_i)|\ i = 0, 1, 2, \ldots, N-1\}$, where $\mathbf{x}_i$ denotes the i_{th} frame with corresponding label of y_i. All the frames and labels are extracted from the raw signal $\mathbf{X}$ and annotation vector $\mathbf{Y}$, respectively, which can be represented by

$$x_i[1, 2, \ldots, l] = X[(1 + s*i), \ldots, (1 + s*i + l)] \quad (1)$$
$$y_i = Y[1 + (s*i) + l/2]$$
$$\forall i \in \{0, 1, 2, \ldots, N-1\} \quad (2)$$

where l denotes the length of each frame, s is the frame shift, and N is the total number of frames. Hence, smaller shift, s, between subsequent frames will increase overlapping samples between frames.

A neural network model will be trained using these extracted frames and corresponding labels. Hence, a binary cross entropy loss function($\mathscr{L}$) can be defined that will be optimized in training stage for generating correct predictions from the neural network, which is given by

$$\mathscr{L}(w) = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \ln(\tilde{y}_i) + (1 - y_i)\ln(1 - \tilde{y}_i)\right]$$
$$+ \frac{\lambda}{2n}\sum_{w} w^2 \quad (3)$$

where n is the total number of samples in a batch, w denotes the weight vector, y_i and $\tilde{y}_i$ are the actual label and generated prediction, respectively, for the i_{th} input, and λ is the regularization parameter adjusted for reducing overfitting.

During the evaluation phase, the trained and optimized model is used to generate predictions about occurrence of apnea in a frame by frame basis. For every case, the frame is labelled according to the annotation of the mid-sample of EEG frame.

B. VARIATIONAL MODE DECOMPOSITION (VMD)

As the apnea events lead to some complicated neurological activities, these may excite the patient in sleep to balance out the abnormalities. Hence, some random variations of spectral contents can be introduced in the EEG signals of different apnea patients. The time domain representation of the EEG signals doesn't provide the effective representation for detecting the apnea events restricting the improvements in recognition accuracy. Neurological activity of the brain changes from non-apnea to apnea episodes and this can cause significant variation in the various frequency bands of EEG signal which are difficult to observe and capture in the raw time domain signal. The dominant frequencies in these frequency bands caused by the neural activities can be expected to shift with time and depending on the person. Such limitations can be overcome by using a dynamic band division with adaptive center frequencies which can be achieved by using variational mode decomposition (VMD) algorithm [22]. Since VMD decomposes EEG into modes whose center frequencies are adaptively determined, these modes can better follow the frequency shifts and offer an improved representation of the apnea related neural variations. Hence, VMD facilitates the optimization process of the deep neural network for converging to the optimal solution by extracting more effective features. Moreover, VMD increases the computational overhead minimally to gain better performance compared to the end-to-end deep learning approaches.

As proposed in [29], VMD decomposes an input signal, $y(t)$, into predefined N number of principal modes, $m_i(t)$, which is given by

$$y(t) = \sum_{i=1}^{N} m_i(t) \quad (4)$$

These modes are dynamically determined to minimize the sum of bandwidths of all modes while reconstructing the input signal at the same time in least squares sense through the addition of modes. Hence, the constrained variational optimization problem becomes

$$\min_{\{m_i\},\{\omega_i\}} : \left\{\sum_{i=1}^{N}\left\|\partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right) * m_i(t)\right]e^{-j\omega_i t}\right\|_2\right\} \quad (5)$$

where m_i and w_i denote the i_{th} mode function and its corresponding center frequency, respectively.

However, several recording environmental conditions along with varying neurophysiological characteristics at

different sleep stages may cause misleading amplitude contrast in-between frames of different subjects. These issues can be overcome by carrying out DC offset removal followed by amplitude normalization at the pre-processing stage. After that, the variational mode decomposition algorithm is incorporated to split each frame into number of modes that are processed later using the proposed neural network.

C. PROPOSED DEEP NEURAL NETWORK

As shown in Fig. 1, the proposed deep neural network architecture consists of three separate sub-networks: a fully convolutional neural network, a bi-directional LSTM network, and densely connected layers. Firstly, each mode of EEG frame obtained from the VMD is processed separately using a fully convolutional neural network (FCNN). Number of such FCNNs are used in parallel to extract the temporal feature variations from each mode separately. These operations transform the feature space of each mode into a comparatively low temporal dimension while maintaining the causal temporal relationships among the extracted features. This process facilitates the extraction of the general trends of unimodal feature variations that are supposed to occur during apnea events irrespective of patients. Though dividing the raw EEG frames into number of modes with adaptive center frequencies facilitates the temporal feature extraction of apnea events, significant inter-modal temporal relationships exist among these modes. Hence, output feature maps obtained from each FCNN module operating with different modes of EEG-frames are subjected to processing together. Later, a multi-layer bi-directional long short term memory network (BiLSTM) is being fed with these temporal feature maps obtained from different FCNNs. This module operating with multi-modal features of EEG frames extract high dimensional inter-modal temporal features that are supposed to vary during apnea events. Eventually a combined feature vector is generated from the BiLSTM module that contains generic feature representation of all the modes of the corresponding EEG frame. This feature vector is processed with series of densely connected layers that extract the general representations among extracted features to converge towards the final prediction of apnea event. Detailed architectural analysis of all these sub-networks is provided in the following discussion.

1) PROPOSED FULLY CONVOLUTIONAL NEURAL NETWORK (FCNN)

All the operations in the proposed FCNN can be divided into two blocks of operations in general, which are: causal dilation block and multi-kernel block (Fig. 2). At any instant of the EEG signal, all the previous history of the EEG signal should be taken into consideration to distinguish the neurological activity pattern that instigates the apnea event. Furthermore, for proper recognition of the apnea event, the variations of the EEG signal should be analyzed over a longer temporal window. As short-term fluctuations at various stages of sleep may lead to improper recognition of apnea events,
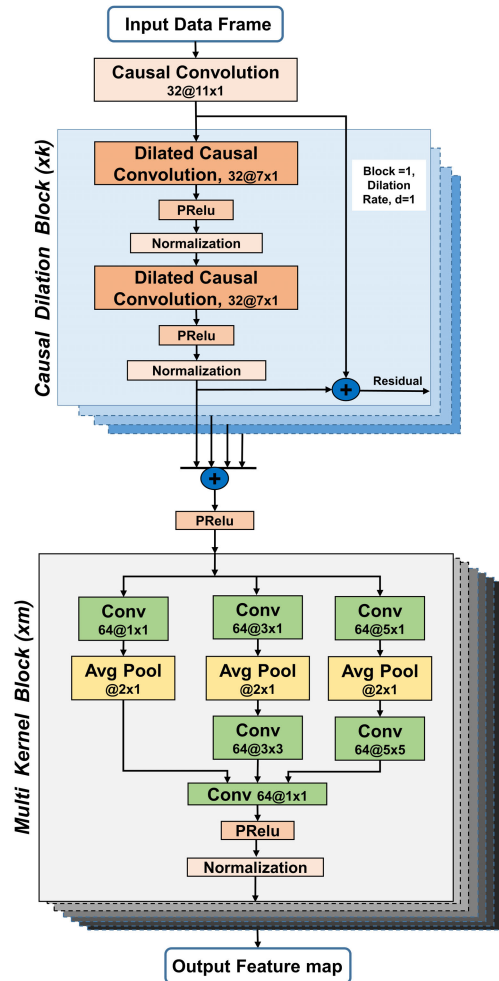

FIGURE 2. Proposed Fully Convolutional Neural Network (FCNN) with **k** number 'Causal dilation block' and 'm' number 'Multi kernel block'. Here, (filter@kernel size) represents the number of filters and kernel size in each convolution.

EEG signals should be analyzed from different range of observations. Both of these objectives are incorporated in the 'Causal Dilation Block' (CDB) utilizing a series of dilated causal convolutions. As causal convolution operation at any time instant takes into consideration all the previous history [30], it achieves the first objective. However, a very deep architecture is needed to achieve a large receptive field by only using causal convolution. Hence, by utilizing dilated convolution, the receptive area can be increased by a large margin (shown in Fig. 3). Causal dilated convolutions are widely used in numerous speech related applications relating speech generation, denoising and many others, where the context is to be extracted from large window [31], [32]. If the input signal corresponds to $x \in \mathbb{R}^n$ and the filter of length k is represented by, $f : \{0, 1, \ldots, k-1\} \to \mathbb{R}$, the causal dilated convolution operation F on any element s is given by

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \, \mathbf{x}_{(s-d)i} \qquad (6)$$
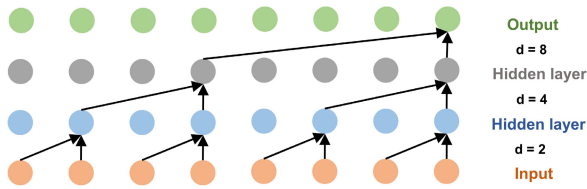
FIGURE 3. Dilated causal convolutions with increasing dilation rates.

where $*_d$ denotes the causal dilated convolution with dilation rate of d and $(s - d)$ represents the causal convolution with previous time stamps. Hence, the effective receptive area in one such layer is $(k - 1)d$. The dilation rate is varied exponentially with $(d = O(2^i))$ at the i-th block.

In each of such block, two causal dilated convolutions are done in series with application of parametric rectified linear unit (PRelu) [33] as nonlinear activation function followed by normalization operation, where

$$\text{PRelu(x)} = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{otherwise} \end{cases} \quad (7)$$

Here, α is the slope used for mapping negative input values. In this study, $\alpha = 0.2$ is chosen for its faster convergence.

If f_i is the output feature map obtained from the i_{th} CDB block, a residual output, R_i, is also generated that will be fed to the following CDB block which can be represented by

$$R_i = f_i + R_{i-1} \quad \forall i = \{1, 2, \ldots, k\} \quad (8)$$

where $R_0 = x_i$ is the input signal and k is the total number of causal dilation blocks. Afterward, a combined output, $\mathcal{F}$, is generated combining the features extracted using each of the CDB unit blocks that can be expressed as

$$\mathcal{F} = \sum_{i=1}^{k} f_i \quad (9)$$

After processing with number of causal dilation blocks, the input dataframe of corresponding EEG mode is transformed to a resultant feature map that contains numerous extracted features containing diverse temporal relationships. However, the temporal dimension of this feature map is kept as same to the input dataframe which should be reduced further to extract the more general unimodal temporal features. Thus, multi kernel block (MKB) is proposed to reduce the temporal dimension of the feature map while performing convolution operations with multiple kernels in parallel. Here, average pooling operation is carried out in between convolutions with different kernels in parallel that helps to incorporate diverse temporal contexts in the pooling operation. Subsequently, all these pooled and convolved feature maps are converged together with another convolution operation that extracts the general variations. Hence, output from each of the MKB units can be generalized as

$$O_i = H(h_1(O_{(i-1)}, \theta_1), h_2(O_{(i-1)}, \theta_2), h_3(O_{(i-1)}, \theta_3), \theta_H)$$
$$\forall i = \{1, 2, \ldots, m\} \quad (10)$$

where $O_0 = \mathcal{F}$; h_1, h_2, h_3 represent convolutions with temporal kernels of size 1, 3 and 5, respectively; H represents the combined convolution while $\theta_1, \theta_2, \theta_3$, and θ_H are their respective parameters, and m is the total number of such MKB units. After passing through series of multi kernel blocks, the transformed feature map will incorporate more and more generalized unimodal features with reduced temporal dimensions. Finally, an output feature map is obtained from the final multi-kernel block that contains the generalized temporal representation of a particular EEG mode.

2) PROPOSED BIDIRECTIONAL LONG SHORT-TERM MEMORY NETWORK (BILSTM)

The extracted generalized sequential representations of the modes of the EEG frame obtained from the FCNN module are processed together with the bidirectional LSTM network (shown in Fig. 4). LSTM units [34] are proven to extract long term temporal dependencies used in numerous sequence processing applications. Generally, processing in traditional LSTM units depend on the output of previous units and information flows in one direction. Its improved variant is the bi-directional LSTM units where two such layers of LSTM memory cells process the sequence simultaneously in opposing directions [35]. In many applications, such operations provide more temporal context over a longer time frame. In the proposed BiLSTM module, two of such bidirectional LSTM layers are stacked together. Generalized variational features of different EEG modes undergo through further processing with these LSTM layers that extract effective features of higher levels considering long term inter-modal temporal dependencies.

Hence, both the first forward and backward pass layer process the feature sequence $\{x_1, x_2, \ldots, x_N\}$ produced by the FCNN blocks simultaneously. Each basic LSTM unit cell is comprised of three special data manipulating structures known as input, forget and output gate as outlined in Fig. 4b. Here, σ and tanh denote the logistic sigmoid and hyperbolic tangent activation functions, respectively. The relations used for calculating hidden state $\overrightarrow{h_t}$ of a forward layer are given as follows:

$$\overrightarrow{i}_t = \sigma(\overrightarrow{W}_{ix} \overrightarrow{x}_t + \overrightarrow{W}_{ih} \overrightarrow{h}_{t-1} + \overrightarrow{b}_i) \quad (11)$$

$$\overrightarrow{f}_t = \sigma(\overrightarrow{W}_{fx} \overrightarrow{x}_t + \overrightarrow{W}_{fh} \overrightarrow{h}_{t-1} + \overrightarrow{b}_f) \quad (12)$$

$$\overrightarrow{\tilde{C}}_t = \tanh(\overrightarrow{W}_{cx} \overrightarrow{x}_t + \overrightarrow{W}_{ch} \overrightarrow{h}_{t-1} + \overrightarrow{b}_c) \quad (13)$$

$$\overrightarrow{C}_t = \overrightarrow{f}_t \odot \overrightarrow{C}_{t-1} + \overrightarrow{i}_t \odot \overrightarrow{\tilde{C}}_t \quad (14)$$

$$\overrightarrow{o}_t = \sigma(W_{ox} x_t + W_{oh} \overrightarrow{h}_{t-1} + b_o) \quad (15)$$

$$\overrightarrow{h}_t = \overrightarrow{o}_t \odot \tanh(\overrightarrow{c}_t) \quad (16)$$

In our stacked model, hidden states of first forward layer and backward layer are concatenated and passed on to both the secondary forward and backward LSTM layers for more processing of inter-modal temporal relations. These secondary hidden layers further explore the feature space to produce a sequence of feature vectors for forward and backward
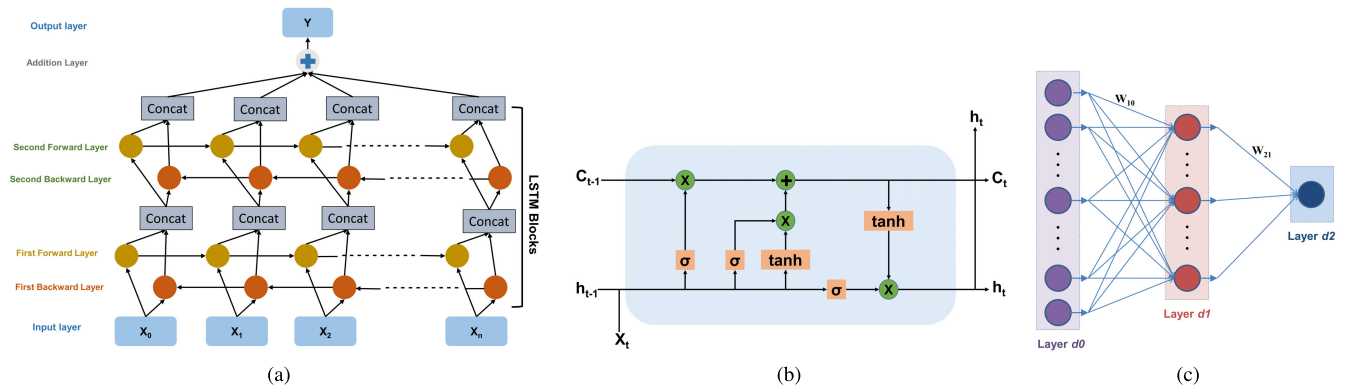
FIGURE 4. Schematic representation of (a) the proposed double layer bidirectional LSTM network, (b) LSTM unit block, and (c) densely connected layers.

pass layers, respectively. Finally, after concatenating corresponding feature vectors obtained from each secondary forward and backward pass LSTM unit, all of these feature vectors are added together to produce the final output feature vector. Therefore, stack of bidirectional LSTM layers converge all the temporal inter-modal features of different modes of respective EEG frame into a resultant feature vector that contains the global temporal representation of that particular frame.

3) DENSELY CONNECTED LAYERS AND CLASSIFIERS
The resultant temporal feature vector is needed to be converged into the final prediction of apnea incident. Series of densely connected layers are used to exploit the global relationships across all the extracted temporal features as shown in Fig. 4c. This can be given by

$$d_i = \sigma(W_i d_{i-1} + \mathbf{b_i}) \quad \forall i \in \{1, 2, 3\} \qquad (17)$$

where d_i represents the output of the i-th densely connected layer with W_i weight matrix and b_i bias vector, σ represents the activation function and d_0 is the input feature vector. In total, three densely connected layers are stacked in series for converging the output feature vector of stacked LSTM units toward the final prediction. Finally, the output obtained from the final densely connected layer with single node is mapped into the final prediction using sigmoid activation function, which is given by

$$\text{sigmoid(x)} = \frac{1}{1 + e^{-x}} \qquad (18)$$

III. RESULTS
In this section, results obtained from extensive experimentation on various publicly available databases will be discussed and analyzed from diverse perspectives.

A. DATABASE
The proposed method was applied on three public datasets to validate the applicability and robustness of the proposed scheme in subject independent scenario.

1) ST. VINCENT's UNIVERSITY HOSPITAL/UNIVERSITY COLLEGE DUBLIN SLEEP APNEA DATABASE [36]
This is a large publicly available database used in a number of other studies. It contains overnight polysomnogram recordings of 4 female and 21 male patients with various Apnea-Hypopnea Index (AHI). The signals were obtained using the Jaeger-Toennies system (Erich Jaeger GmbH, Germany). EEG recordings were collected at 128 Hz. Starting and ending of all the apnea events are manually annotated in the polysomnogram recordings by sleep specialists according to standard scoring rules [37]. Although the dataset contains various common physiological signals for every patient including EOG, EMG, and EEG, we have focused solely on EEG for its ease of collection without disturbing the patient during sleep, and decomposed the signal after dc offset removal and normalization of the data.

2) YOU SNOOZE YOU WIN - THE PhysioNet COMPUTING IN CARDIOLOGY CHALLENGE 2018 [38]
Data for this challenge were contributed by the Massachusetts General Hospital's (MGH) Computational Clinical Neurophysiology Laboratory (CCNL), and the Clinical Data Animation Laboratory (CDAC). The sleep stages of the subjects were annotated by clinical staff at the MGH according to the American Academy of Sleep Medicine (AASM) manual. In line with the objective of this paper, only the EEG signals were extracted and pre-processed by removing DC offset and normalizing. In this database, the collected EEG signals were sampled at a rate of 200 Hz.

3) MIT-BIH POLYSOMNOGRAPHIC DATABASE [39]
This database is a collection of recordings of multiple physiologic signals during sleep. Subjects were monitored in Boston's Beth Israel Hospital Sleep Laboratory for evaluation of chronic obstructive sleep apnea syndrome. It contains over 80 hours' worth of polysomnographic recordings, each with an ECG signal annotated beat-by-beat, and EEG and respiration signals annotated with respect to sleep stages and apnea.

For our study, only the EEG signals were extracted from the dataset which had been digitized at a sampling rate of 250 Hz.

In this study, all the annotated apnea-hypopnea events are considered as apnea. The extensive size and wide range of AHI observed among the patients recorded in all the above-mentioned datasets provide adequate opportunity for extensive experimentation.

B. EXPERIMENTAL SETUP

Different hyper-parameters of the network are chosen through experimentation for better performance. Adam optimizer is employed for optimization of the network during training phase. Intel® Xeon® $D - 1653N$ CPU @2.80GHz with 12M Cache and 8 cores along with 48 GB RAM is used for experimentation. For hardware acceleration, 1x NVIDIA RTX 2080 Ti GPU having with 4608 CUDA cores running 1770 MHz with 24 GB GDDR6 memory is deployed.

Subject-independent k-fold cross-validation scheme is employed for evaluation of performance measures. In this scheme, total number of patients is divided into k-subfolds. Hence, in a single stage, all the patients in $(k - 1)$ subfolds are considered for training while the patients in the remaining subfold are used for evaluation by applying the optimum model for the subfold. This process is repeated for k times such that each patient is considered in one of the test-folds and finally, the performance measures are averaged.

C. PERFORMANCE EVALUATION

Traditional metrics are used for evaluating the performance of the proposed scheme, such as accuracy, sensitivity and specificity as described in equations below:

$$\text{Accuracy (Acc.)} = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (19)$$

$$\text{Sensitivity (Sen.)} = \frac{TP}{TP + FN} \times 100 \quad (20)$$

$$\text{Specificity (Spec.)} = \frac{TN}{TN + FP} \times 100 \quad (21)$$

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (22)$$

Since a number of empirical parameters are used in the proposed scheme, all the experimentations are carried out in a systematic way to achieve the optimum value of different parameters. The raw EEG signals are divided into number of frames with uniform length and overlapping in between subsequent frames. With an increase of frame length, the EEG frame contains information over a larger interval that facilitates the apnea detection process though it increases computational complexity along with. Taking both issues into consideration and depending on the available ground truth, the proposed method employs a relatively large frame length of 10 seconds for database [36], [38] and 30 seconds for database [39].

In the subject-independent cross-validation scheme, some frames of different patients are considered in the training and evaluation fold. Results obtained from different

cross-validation schemes with the varying number of folds are summarized in Table 1. With an increasing number of folds, more data are available for a single training stage that provides better performance. It should be noticed that the best performance is achieved in the leave-one-out cross-validation scheme. However, considerable performance is achieved even with a 3-fold cross-validation scheme that contains a significantly smaller number of data for each training stage compared to the best performing leave-one-out scheme. This proves the robustness of the scheme that can learn the representations of apnea events using a significantly smaller amount of training data.

TABLE 1. Performance of the proposed method in different subject-independent cross-validation schemes.

K-fold	Accuracy	Sensitivity	Specificity	F1 Score
3 fold	90.04	89.29	90.84	0.9020
7 fold	91.01	91.19	89.95	0.9086
Leave one out	**93.22**	**91.71**	**93.79**	**0.9276**

Each frame of EEG data is decomposed into a number of mode-functions that are processed in parallel and the extracted features from each mode are jointly optimized later. In Table 2, the effect of the traditional bandpass filtering decomposition method of EEG data is compared with the VMD along with the no decomposition of raw data. The results reported in the first row of Table 2 correspond to the proposed end-to-end DL architecture without utilizing the VMD, which is denoted as Proposed Method (only end-to-end). The results reported in the second and third rows of Table 2 correspond to the combination of the proposed end-to-end DL network with the bandpass filtering decomposition and the VMD, respectively, denoted as Proposed Method (end-to-end with bandpass) and Proposed Method (end-to-end with VMD). It should be noted that adaptive VMD provides higher average accuracy compared to the traditional scheme of band-pass filtering and no-decomposition, respectively, when these are integrated with the proposed scheme. Such improvements signify the effectiveness of the use of VMD in combination with the traditional end-to-end deep learning approaches. As the VMD adaptively adjusts the mode center frequency, it provides more opportunity to extract the general feature variations in the EEG signal introduced for the apnea events in the patient-independent scenario.

TABLE 2. Effect of different eeg band decomposition schemes on the proposed method.

Proposed Method	Acc.	Sen.	Spec.	F1 Score
Only end-to-end	91.83	89.22	91.85	0.9060
End-to-end with Band Pass	90.2	89.45	86.65	0.8850
End-to-end with VMD	**93.22**	**91.71**	**93.79**	**0.9276**

In Fig. 5, the effect of the number of variational mode functions (VMFs) that are used in the VMD algorithm is shown. It is clearly visible that with the increasing number of mode functions, the proposed scheme seems to perform

TABLE 3. Subject-specific cross-validation performance obtained using proposed method (database- [36]).

| Patient ID	Accuracy	Sensitivity	Specificity	F1 Score
UCDDB002	92.79	93.59	91.85	0.93397
UCDDB003	92.34	96.67	87.62	0.92925
UCDDB005	94.92	92.43	97.51	0.94888
UCDDB006	92.77	94.27	90.53	0.93986
UCDDB007	90.6	92.58	89.33	0.88503
UCDDB008	95.06	91.67	98.72	0.95062
UCDDB009	93.77	83.41	98.94	0.89911
UCDDB010	96.31	97.51	87.11	0.97908
UCDDB011	88.81	87.18	89.17	0.73913
UCDDB012	93.14	85.78	99.05	0.91763
UCDDB013	98.52	96.70	99.76	0.98150
UCDDB015	96.29	90.50	99.12	0.94118
UCDDB017	96.82	96.61	97.00	0.96614
UCDDB019	94.08	87.64	99.20	0.92914
UCDDB020	88.46	89.89	86.57	0.89842
UCDDB021	90.27	85.78	96.41	0.91057
UCDDB022	95.26	94.76	95.58	0.94090
UCDDB023	93.82	98.14	90.32	0.93441
UCDDB024	94.70	97.88	91.02	0.95196
UCDDB026	93.28	88.44	97.90	0.92779
UCDDB028	85.52	84.39	86.78	0.86014
Average	**93.22**	**91.71**	**93.79**	**0.9276**
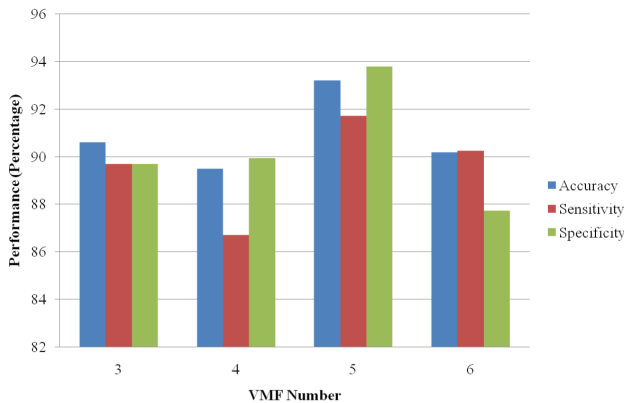
FIGURE 5. Variation of performance for different number of variational mode functions (VMFs) in the proposed scheme.

better and the best performance is attained with 5 modes. For higher number of modes, performance gradually decreases for introducing frequent variations in mode-frequencies which make it difficult to adapt in the patient-independent scenario.

Performance of the proposed scheme on different subjects in the subject-independent leave-one-out cross-validation scheme for databases [36], [38] and [39] are provided in Tables 3, 4 and 5, respectively. After close observation of all the performance metrics obtained for different subjects, it is to be noted that the proposed scheme provides consistent performance irrespective of the subjects. It further proves the robustness of the method that provides significant performance even for unknown subjects in evaluation-phase. Moreover, all the evaluation metrics provide comparable performance that represent the balanced performance of this scheme in both apnea and non-apnea frames.

TABLE 4. Subject-specific cross-validation performance obtained using proposed method (database- [38]).

Patient ID	Accuracy	Sensitivity	Specificity	F1 Score
tr03-0005	87.77	68.42	99.75	0.81
tr03-0029	98.62	90.00	99.33	0.91
tr03-0078	94.81	91.50	97.53	0.94
tr03-0079	89.02	95.89	83.81	0.88
tr03-0083	94.83	94.20	96.68	0.96
tr03-0086	93.35	90.35	95.74	0.92
tr03-0087	92.75	79.55	96.64	0.83
tr03-0092	91.73	98.20	72.13	0.95
tr03-0100	93.66	68.12	98.95	0.79
tr03-0103	97.92	100.00	97.64	0.92
tr03-0134	90.41	99.35	80.34	0.92
tr03-0135	89.90	84.48	96.64	0.90
tr03-0141	89.95	85.86	95.73	0.91
tr03-0146	87.90	95.48	83.21	0.86
tr03-0152	98.13	97.93	98.44	0.98
tr03-0166	95.07	91.83	98.44	0.95
tr03-0167	94.49	97.33	79.03	0.97
tr03-0179	95.81	98.90	93.76	0.95
tr03-0184	89.32	95.94	83.64	0.89
tr03-0187	90.99	88.03	93.52	0.90
tr03-0212	95.76	99.04	71.43	0.98
tr03-0216	97.89	97.71	98.30	0.98
tr03-0241	89.53	92.65	80.49	0.93
tr03-0251	97.73	98.45	97.07	0.98
tr03-0256	91.52	48.51	98.82	0.62
tr03-0257	93.30	94.12	90.07	0.96
tr03-0267	95.64	95.59	95.70	0.96
Average	**93.25**	**90.28**	**91.59**	**0.91**

TABLE 5. Subject-specific cross-validation performance obtained using proposed method (database- [39]).

Patient ID	Accuracy	Sensitivity	Specificity	F1 Score
slp01	95.04	83.78	100.00	0.91
slp02	89.66	83.53	98.33	0.90
slp03	90.38	92.27	86.08	0.93
slp04	96.88	97.28	95.79	0.98
slp14	94.67	96.20	93.33	0.94
slp16	87.91	93.40	77.98	0.91
slp32	82.78	86.67	77.05	0.86
slp37	99.08	100.00	98.09	0.99
slp48	86.71	89.50	78.21	0.91
slp59	89.06	97.26	63.04	0.93
slp60	92.36	85.42	95.41	0.87
slp61	74.60	61.38	93.20	0.74
slp66	90.25	89.60	91.35	0.92
slp67	82.35	93.10	74.36	0.82
Average	**89.41**	**89.24**	**87.30**	**0.90**

Stack of bidirectional LSTM layers is introduced for further processing of the temporal features obtained from the FCNN module with different modes. In Table 6, the effect of different number of nodes in the LSTM units of each bidirectional LSTM layer is investigated and summarized. The model provides optimum performance in most cases with 128 and 256 nodes in the LSTM units of first and second Bi-LSTM layers, respectively. With more number of nodes, the LSTM units employ more memory units that can extract longer temporal dependencies of inter-modal features. However, with a large number of nodes, the network gets more intricate as well which makes the convergence difficult for vanishing gradient and overfitting issues.

TABLE 6. Effect of different number of nodes in the stack of bidirectional LSTM layers.

Node Number	Accuracy	Sensitivity	Specificity	F1 Score
[128, 256]	**93.22**	91.71	**93.79**	**0.9276**
[128, 128]	92.89	**92.53**	92.37	0.9264
[256, 256]	92.60	90.51	92.36	0.9154

TABLE 8. Comparison with other existing approaches.

Methods	Average Accuracy(%)	Average Sensitivity(%)	Average Specificity(%)
Zhou et al. [19]	59.98	55.39	65.21
Wafaa et al. [41]	67.92	51.18	81.77
Bhattacharjee et al. [22]	80.18	85.38	74.19
Tanvir et al. [40]	88.22	85.52	88.97
Proposed (end to end)	91.83	89.22	91.85
Proposed	**93.22**	**91.71**	**93.79**

After complete training and optimization, apnea frames are predicted for each unknown subject. The conditions affecting the quality of collected EEG signal may be quite different for each sleep center and the staff administering the process introducing difficulty in apnea prediction. Furthermore, naturally each subject exhibits unique biological and physiological characteristics leading to variations in the phenomena experienced during sleep apnea events. These changes may not be properly reflected in networks trained on other patients. Incorporating these patient specific information into the scheme can help the network to adapt and tailor its response to better accommodate each subject. Though in clinical scenario this would require manual annotation of a segment of the data, this can still reduce the burden on professionals and speed up the diagnosis process significantly. To explore this effect, a semi-supervised approach is explored where in the training phase, 30% and 50% of the data of the subject under consideration is included. As illustrated in Table 7, the insights gained from the new data boosts the prediction capabilities of the network. Utilizing a larger portion of the newly recruited subject's data is not feasible.

TABLE 7. Effect of overlap of test subjects in the training set.

Overlap	Accuracy	Sensitivity	Specificity	F1 Score
0%	93.22	91.71	93.79	0.9276
30%	94.88	94.53	93.19	0.9406
50%	**95.54**	**95.35**	**93.00**	**0.9440**

In Table 8, the proposed scheme is compared with other state-of-the-art methods of automated apnea detection using singular channel EEG based schemes. While considering the comparisons, it must be noted that most of these approaches reported performance in a subject-dependent manner keeping data from same subject in training and testing set. These approaches are generally based on handcrafted feature extraction process along with traditional shallow classifier that makes the apnea detection task very difficult in a subject-independent scenario. As these features undergo significant variations with random perturbations of the high-frequency EEG data due to noises and other artifacts, such hand-crafted features aren't enough to extract the general feature variations that are introduced for apnea events. Bhattacharjee *et al.* used traditional band division of EEG data [21] as well as VMD [22] with rician modeling of entropy and log-variance features and proved the precedence of VMD of EEG signal while detecting apnea. However, it is to be noted that such handcrafted feature extractors aren't enough to provide considerable performance. Zhou *et al.* [19]

introduced a unique method to predict sleep apnea by analysing detrended fluctuation for feature extraction. This unique method also suffered from similar limitations. In our previous work [40], an end-to-end fully connected convolutional network incorporating residual units was proposed to automate the feature extraction process using full band EEG data. By using the end-to-end network, it provides better performance than [19], [41] and [22], thereby proves the advantage of deep learning in the prediction of sleep apnea. However it has lack of architectural variations in the CNN module to process the long term temporal variations of apnea events in EEG data along with the additional complexity of operating with EEG data in raw form. Due to the application of more sophisticated architectural blocks along with Bi-LSTM modules, the proposed end-to-end DL network outperforms the previous approaches even without using the VMD. However, It is clearly observed that the proposed method (end-to-end DL with VMD) provides the best achievable performance in all of the evaluation metrics compared to the other approaches reported in the table. To be precise, the sensitivity has been improved by more than 2% which is significant for any disease diagnosis scheme.

IV. DISCUSSION

In this study, an automated sleep apnea frame detection scheme is proposed using a deep fully convolutional-BiLSTM neural network for the subject-independent test scenario. It outperforms the existing apneic frame detection approaches significantly in all performance metrics using only single channel EEG signal. Consistently high accuracy and F1 scores are obtained for all three databases which have different properties such as sampling rate, data collection procedures and scoring standards. Two major reasons for getting very satisfactory performances are (1) utilizing the VMD operated input data and (2) employing the proposed end-to-end deep learning network. Instead of directly using the raw EEG data, the variational mode decomposed input EEG data are used. The advantages of using variational mode decomposed EEG signals in an end-to-end deep learning network are summarized as follows:

- improved performance with minimal computational overhead,
- better representation of the apnea events for gaining better generalization with the deep neural network, and
- improving explainability of the achieved performance.

According to Table 8, it is observed that the use of VMD in combination with the proposed end-to-end deep learning

network offers relatively better performance compared to the other methods. Such improvements signify the effectiveness of the use of VMD in combination with the traditional end-to-end deep learning approaches. Moreover, VMD increases the computational overhead minimally to gain better performance compared to the end-to-end deep learning approaches. Generally, neural activity varies during non-apnea and apnea periods, which is reflected in different EEG frequency bands. These changes are more distinct in the variational mode decomposed signals as the center frequencies of the modes in the VMD are adaptively calculated, it allows to capture the center frequency shift due to variation of neurological activity and reduces the burden on the proposed end-to-end deep neural network model to learn the complex relationships on its own. Hence, VMD facilitates the optimization process of the deep neural network for converging to the optimal solution by extracting more effective features. Furthermore, VMD facilitates the interpretability of the achieved performance by extracting the effects of the different frequency modes in the overall recognition performance. It can be observed from the results that the inter-modal information learned from the decomposed EEG modes generated through VMD helped improve the classification performance. The statistical significance of the improvement is proven by performing a paired T-test where a p-value of 0.0329 is obtained.

Use of dilated causal convolutions with the LSTM modules helps the proposed end-to-end deep learning network extract the general pattern of EEG variations for apnea events. As explained earlier, the stack of bidirectional LSTM layers offer a global temporal representation of a particular frame utilizing the temporal inter-modal features of different modes of respective EEG frame. However, further improvement of the proposed apnea detection scheme can be achieved by incorporating effective learning from other relevant applications into the proposed scheme using non-conventional learning mechanism called 'meta-learning' and other similar transfer learning based schemes. The basic idea of meta-learning is to effectively transfer knowledge from several similar tasks for extracting the general representation to learn a new task with fewer training samples. In [42], this mechanism was used for sleep stage classification task by adopting a transfer learning based technique to utilize sleep staging knowledge acquired from a large dataset to classify frames of unseen new subjects. As such knowledge-transfer based scheme from relevant other applications is beyond the scope of this study, it is left as a potential future work. It is to be noted that a major limitation inherent to most of the sleep apnea detection methods using EEG signal is the dependency on availability of a large amount of data of any new subject. In clinical scenario, when diagnosing a new subject, there will not be pre-existing annotated frames of such large volume rendering the methods infeasible and impractical for real life applications. Only a patient independent scheme where a network is trained on data of already diagnosed patients can be employed in real life situation. In addition, as can be seen in the semi-supervised approach of this paper and as

evidenced by the results obtained in [43], such patient dependent approaches can produce misleading overestimation of the prediction accuracy. The proposed subject independent method does not suffer from these drawbacks and is suitable for practical deployment.

V. LIMITATIONS OF THE STUDY

While the results for individual datasets show promising performance, there are some limitations that need to be acknowledged. Training the proposed model requires a significant number of normal and apneic EEG frames from various patients. Reduction in the number of frames in the training set results in slight declination in the performance, as can be seen by comparing the patient independent cross validation results of dataset- [39] and dataset- [36]. As the number of patients in [36] was higher than the latter, the model seems to perform better for this dataset than dataset- [39]. However, this might not be a problem anymore due to the easy access to different public datasets. Another minor issue to be acknowledged is the absence of transfer learning from relevant other applications in the proposed model, as our goal was to analyze the advantage of using variational mode decomposed EEG data on an end-to-end deep learning network, while restricting the use of bio-signals to only one type (EEG signal). Lastly, our current study only considers the binary classification between sleep apnea frames and normal EEG frames without detecting the type of apnea. We have left the detection of multiclass apnea frames as well the addition of meta-learning or other transfer learning based approaches in the proposed method for future studies.

VI. CONCLUSION

In this paper, an automated sleep apnea frame detection scheme is proposed using deep fully convolutional-BiLSTM neural network for the subject-independent test scenario. For efficient processing of raw EEG data, variational mode decomposition is adopted for introducing adaptive variations in center mode frequencies. It is shown that such decomposition contributes considerably towards extracting the subject independent feature variations for apnea events resulting in substantial performance improvements compared to the raw EEG data processing. Furthermore, VMD facilitates the interpretability of the achieved performance by extracting the effects of the different frequency modes in the overall recognition performance. To automate the process of feature extraction as well as to analyze the long term temporal variations in the EEG data, the proposed network utilizes dilated causal convolutions with long-short term memory modules effectively that is proven to extract the general pattern of EEG variations for apneic episodes irrespective of subjects. Through experimentation, individual parameters of the proposed method are varied and optimal values are found that provide satisfactory prediction performance using the proposed scheme. Consistent performance is achieved for all the individual subjects in subject independent cross validation scheme. Moreover, for achieving further efficiency

in the prediction of unknown EEG frames, a semi-supervised algorithm is introduced that increases the apnea detection performance by leveraging insight gained about the individual subject's conditions. Despite using separate subjects in the training and evaluation phase, significantly higher performance is achieved using the proposed scheme in all the evaluation metrics that ensures the applicability and robustness of this method for practical applications of sleep apnea occurrence predictions.

REFERENCES

[1] T. E. Weaver and C. F. George, "Cognition and performance in patients with obstructive sleep apnea," in *Principles and Practice of Sleep Medicine*, 4th ed. Amsterdam, The Netherlands: Elsevier, 2005, ch. 85, pp. 1023–1033.

[2] P. E. Peppard, T. Young, J. H. Barnet, M. E. W. P. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *Amer. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, May 2013.

[3] P. E. Peppard, T. Young, M. Palta, and J. Skatrud, "Prospective study of the association between sleep-disordered breathing and hypertension," *England J. Med.*, vol. 342, pp. 1378–1384, May 2000.

[4] V. K. Somers, D. P. White, R. Amin, W. T. Abraham, F. Costa, A. Culebras, S. Daniels, J. S. Floras, C. E. Hunt, L. J. Olson, T. G. Pickering, R. Russell, M. Woo, and T. Young, "Sleep apnea and cardiovascular disease," *Circulat.*, vol. 118, no. 10, pp. 1080–1111, 2008.

[5] K. Yaffe, A. M. Laffan, S. L. Harrison, S. Redline, A. P. Spira, K. E. Ensrud, S. Ancoli-Israel, and K. L. Stone, "Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women," *JAMA*, vol. 306, no. 6, pp. 613–619, Aug. 2011.

[6] H. C. Kim, T. Young, C. G. Matthews, S. M. Weber, A. R. Woodard, and M. Palta, "Sleep-disordered breathing and neuropsychological deficits. A population-based study," *Amer. J. Respiratory Crit. Care Med.*, vol. 156, no. 6, pp. 1813–1819, 1997.

[7] P. E. Peppard, M. Szklo-Coxe, K. M. Hla, and T. Young, "Longitudinal association of sleep-related breathing disorder and depression," *Arch. Internal Med.*, vol. 166, no. 16, pp. 1709–1715, 2006.

[8] N. S. Marshall, K. K. Wong, P. Y. Liu, S. R. Cullen, M. W. Knuiman, and R. R. Grunstein, "Sleep apnea as an independent risk factor for all-cause mortality: The Busselton health study," *Sleep*, vol. 31, no. 8, pp. 1079–1085, 2008.

[9] F. Karimzadeh, M. Nami, and R. Boostani, "Sleep microstructure dynamics and neurocognitive performance in obstructive sleep apnea syndrome patients," *J. Integrative Neurosci.*, vol. 16, no. 2, pp. 127–142, Feb. 2018.

[10] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to diagnosis and treatment of patients with suspected sleep apnea," *Amer. J. Respiratory Crit. Care Med.*, vol. 169, no. 6, pp. 668–672, Mar. 2004.

[11] N. A. Collop, "Scoring variability between polysomnography technologists in different sleep laboratories," *Sleep Med.*, vol. 3, no. 1, pp. 43–47, 2002.

[12] D. Bliwise, N. G. Bliwise, H. C. Kraemer, and W. Dement, "Measurement error in visually scored electrophysiological data: Respiration during sleep," *J. Neurosci. Methods*, vol. 12, no. 1, pp. 49–56, Nov. 1984.

[13] M. Al-Angari and V. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 463–468, May 2012.

[14] J. A. Waxman, D. Graupe, and D. W. Carley, "Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network," *Amer. J. Respiratory Crit. Care Med.*, vol. 181, no. 7, pp. 727–733, Apr. 2010.

[15] D. Alvarez, R. Hornero, J. V. Marcos, F. Del Campo, and M. Lopez, "Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 400–403.

[16] X. Chen and Z. J. Wang, "Design and implementation of a wearable, wireless EEG recording system," in *Proc. 5th Int. Conf. Bioinf. Biomed. Eng.*, May 2011, pp. 1–5, doi: 10.1109/icbbe.2011.5781501.

[17] A. R. See and C.-K. Liang, "A study on sleep EEG using sample entropy and power spectrum analysis," in *Proc. Defense Sci. Res. Conf. Expo (DSR)*, Aug. 2011, pp. 1–4, doi: 10.1109/dsr.2011.6026802.

[18] 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.

[19] J. Zhou, X.-M. Wu, and W.-J. Zeng, "Automatic detection of sleep apnea based on EEG detrended fluctuation analysis and support vector machine," *J. Clin. Monitor. comput.*, vol. 29, no. 6, pp. 767–772, 2015.

[20] M. F. Aksahin, A. Erdamar, A. Isik, and A. Karaduman, "Sleep apnea detection using with EEG, ECG and respiratory signals," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4, doi: 10.1109/siu.2017.7960210.

[21] A. Bhattacharjee, S. Saha, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Sleep apnea detection based on rician modeling of feature variation in multiband EEG signal," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1066–1074, May 2019.

[22] A. Bhattacharjee, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "VMD-RiM: Rician modeling of temporal feature variation extracted from variational mode decomposed eeg signal for automatic sleep apnea detection," *IEEE Access*, vol. 6, pp. 77440–77453, 2018.

[23] E. Svanborg and C. Guilleminault, "EEG frequency changes during sleep apneas," *Sleep*, vol. 19, no. 3, pp. 248–254, 1996.

[24] H. P. R. Bandla and D. Gozal, "Dynamic changes in EEG spectra during obstructive apnea in children," *Pediatric Pulmonol.*, vol. 29, no. 5, pp. 359–365, May 2000.

[25] C.-C. Hsu and P.-T. Shih, "A novel sleep apnea detection system in electroencephalogram using frequency variation," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6014–6024, 2011.

[26] S. Saha, A. Bhattacharjee, and S. A. Fattah, "Automatic detection of sleep apnea events based on inter-band energy ratio obtained from multi-band EEG signal," *Healthcare Technol. Lett.*, vol. 6, no. 3, pp. 82–86, Jun. 2019.

[27] J. M. Kang, S. T. Kim, S. Mariani, S.-E. Cho, J. W. Winkelman, K. H. Park, and S.-G. Kang, "Difference in spectral power density of sleep EEG between patients with simple snoring and those with obstructive sleep apnoea," *Sci. Rep.*, vol. 10, no. 1, p. 6135, Dec. 2020, doi: 10.1038/s41598-020-62915-x.

[28] C. Shahnaz, A. T. Minhaz, and S. T. Ahamed, "Sub-frame based apnea detection exploiting delta band power ratio extracted from EEG signals," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 190–193.

[29] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[30] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*. [Online]. Available: http://arxiv.org/abs/1609.03499

[31] Y. Luo and N. Mesgarani, "Conv-TasNet: Surpassing ideal time–frequency magnitude masking for speech separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 8, pp. 1256–1266, Aug. 2019.

[32] D. Rethage, J. Pons, and X. Serra, "A wavenet for speech denoising," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5069–5073.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[35] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 4, Aug. 2005, pp. 2047–2052.

[36] W. Nicholas, L. Doherty, S. Ryan, J. Garvey, P. Boyle, and E. Chua. (2004). *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*. physionet.org. [Online]. Available: https://physionet.org/content/ucddb/

[37] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stage of Human Subjects*. Washington, DC, USA: Public Health Service U.S. Gov. Printing Office, 1968.

[38] M. Ghassemi, B. Moody, L.-W. Lehman, C. Song, Q. Li, H. Sun, B. Westover, and G. Clifford, "You snooze, you win: The PhysioNet/computing in cardiology challenge 2018," in *Proc. Comput. Cardiol. Conf. (CinC)*, vol. 45, Dec. 2018, pp. 1–4, doi: 10.22489/cinc.2018.049.

[39] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry Clin. Neurosci.*, vol. 53, no. 2, pp. 175–177, 1999.

[40] T. Mahmud, I. A. Khan, T. Ibn Mahmud, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Sleep apnea event detection from sub-frame based feature variation in EEG signal using deep convolutional neural network," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 5580–5583.

[41] W. S. Almuhammadi, K. A. Aboalayon, and M. Faezipour, "Efficient obstructive sleep apnea classification based on EEG signals," in *Proc. Long Island Syst., Appl. Technol.*, May 2015, pp. 1–6, doi: 10.1109/LISAT.2015.7160186.

[42] N. Banluesombatkul, P. Ouppaphan, P. Leelaarporn, P. Lakhan, B. Chaitusaney, N. Jaimchariyatam, E. Chuangsuwanich, W. Chen, H. Phan, N. Dilokthanakul, and T. Wilaiprasitporn, "MetaSleepLearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1949–1963, Jun. 2021.

[43] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *Giga-science*, vol. 6, no. 5, May 2017, Art. no. gix019, doi: 10.1093/giga-science/gix019.

SHAIKH ANOWARUL FATTAH (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, and the Ph.D. degree in electronics and communication engineering (ECE) from Concordia University, Canada. He held a visiting postdoctoral position at Princeton University, NJ, USA. He is currently serving as a Professor for the Department of Electrical and Electronics Engineering (EEE), BUET. He has published more than 215 international journal articles/conference papers with some best paper awards and delivered more than 80 Keynote/invited talks in many countries. His research interests include signal processing, machine learning, and biomedical engineering. He is a fellow of IEB. He is a member of IEEE PES LRP, the IEEE Public Visibility Committee, and the IEEE Smart Village Education Committee. He received several awards, namely the Concordia University's Distinguished Doctoral Dissertation Prize in ENS, in 2009, the 2007 URSI Canadian Young Scientist Award, the Dr. Rashid Gold Medal (in M.Sc.), the BAS-TWAS Young Scientists Prize, in 2014, the 2016 IEEE MGA Achievement Award, the 2017 IEEE R10 HTA Outstanding Volunteer Award, and the 2018 IEEE R10 Outstanding Volunteer Award. He served as the Chair/the Founding Chair for different IEEE Society chapters in Bangladesh, such as IEEE SPS, EMBS, RAS, and SSIT. He was the IEEE Bangladesh Section Chair for the period 2015–2016. He is the Chair of IEEE PES-HAC. He served on IEEE HAC and various committees for IEEE R10, IEEE EAB, and IEEE SIGHT. He served key positions in many international conferences, such as the General Chair of IEEE R10-HTC2017 and the TPC Chair of IEEE TENSYMP2020. He is the Editorial Board Member of IEEE Access and *BioMed Research International*, and an Editor of IEEE PES Enews, *JEE*, and *IEB*.

TANVIR MAHMUD received the B.Sc. and M.Sc. degrees from the Department of Electrical and Electronics Engineering (EEE), Bangladesh University of Engineering and Technology. He is currently serving as a Lecturer for Department of EEE, Bangladesh University of Engineering and Technology. His research interests include computer vision, hardware acceleration, biomedical signal processing, image processing, and machine learning.

ISHTIAQUE AHMED KHAN received the B.Sc. degree from the Department of Electrical and Electronics Engineering (EEE), Bangladesh University of Engineering and Technology, where he is currently pursuing the master's degree. His research interests include signal processing, deep learning, image processing, biomedical signal processing, and machine learning.

WEI-PING ZHU (Senior Member, IEEE) received the B.E. and M.E. degrees from the Nanjing University of Posts and Telecommunications, and the Ph.D. degree from Southeast University, Nanjing, China, in 1982, 1985, and 1991, respectively, all in electrical engineering.

From 1991 to 1992, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada, where he was also a Research Associate, from 1996 to 1998. From 1993 to 1996, he was an Associate Professor with the Department of Information Engineering, Nanjing University of Posts and Telecommunications. From 1998 to 2001, he was with Hi-Tech companies in Ottawa, Canada, including Nortel Networks and SR Telecom Inc. Since 2001, he has been with the Department of Electrical and Computer Engineering, Concordia University, as a full-time Faculty Member, where he is currently a Full Professor. His research interests include digital signal processing fundamentals, speech and statistical signal processing, and signal processing for wireless communication, with a particular focus on MIMO systems and cooperative communication. He was the Chair-Elect of the Digital Signal Processing Technical Committee (DSPTC), IEEE Circuits and System Society, from 2012 to 2014, and the Chair of the DSPTC, from 2014 to 2016. He has served as an Associate Editor for the IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, from 2001 to 2003, an Associate Editor for *Circuits, Systems and Signal Processing*, from 2006 to 2009, and IEEE Transactions on Circuits and Systems—II: Express Briefs, from 2011 to 2015. He was also a Guest Editor of the IEEE Journal on Selected Areas in Communications for the Special Issues of Broadband Wireless Communications for High Speed Vehicles, and Virtual MIMO, from 2011 to 2013. He is an Associate Editor of the *Journal of The Franklin Institute*.

TALHA IBN MAHMUD received the B.Sc. degree from the Department of Electrical and Electronics Engineering (EEE), Bangladesh University of Engineering and Technology, where he is currently pursuing the master's degree. He is currently serving as a Lecturer for the Department of EEE, Green University of Bangladesh. His research interests include signal processing, deep learning, image processing, biomedical engineering, artificial intelligence, and nanomaterials.

M. OMAIR AHMAD (Life Fellow, IEEE) received the B.Eng. degree from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree from Concordia University, Montreal, both in electrical engineering.

From 1978 to 1979, he was a Faculty Member of the New York University College, Buffalo, NY, USA. In 1979, he joined the Faculty of Concordia University, as an Assistant Professor of computer science. He joined the Department of Electrical and Computer Engineering, Concordia University, where he was the Chair of the Department of Electrical and Computer Engineering, from 2002 to 2005. He was an Examiner of the Order of Engineers of Quebec. He was the Founding Researcher at Micronet from its inception, in 1990, as the Canadian Network of Centers of Excellence, until its expiration, in 2004. He is currently a Professor with Concordia University. He holds the Concordia University Research Chair (Tier I) in multimedia signal processing. He has published extensively in the area of signal processing and holds four patents. His current research interests include the areas of multidimensional filter design, speech, image, and video processing, nonlinear signal processing, communication DSP, artificial neural networks, and VLSI circuits for signal processing. In 1988, he was a member of the Admission and Advancement Committee of the IEEE. He was a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation, an induction to Provost's Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science of Concordia University. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. He has served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He was the General Co-Chair of the 2008 IEEE International Conference on Neural Networks and Signal Processing. He is the Chair of the Montreal Chapter IEEE Circuits and Systems Society. He was an Associate Editor of the IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, from 1999 to 2001.

● ● ●