

Exploring Rhythms and Channels-Based EEG Biomarkers for Early Detection of Alzheimer's Disease

Siuly Siuly^{ID}, Member, IEEE, Ömer Faruk Alçin^{ID}, Hua Wang^{ID}, Senior Member, IEEE,
Yan Li^{ID}, Senior Member, IEEE, and Peng Wen^{ID}, Senior Member, IEEE

Abstract—There is no treatment that permanently cures Alzheimer's disease (AD); however, early detection can alleviate the severe effects of the disease. To support early detection of the different stages of AD (e.g., mild, moderate), the key aim of this study is to develop a computer aided diagnostic (CAD) framework that include a long short-term memory (LSTM) network using massive multi-channel electroencephalogram (EEG) data. Although EEG rhythms and EEG channels jointly possess important biomarkers that may be used for diagnosis of AD, but the traditional methods did not explore this issue in any research. To address this problem, this study introduces a new framework to identify the optimal EEG rhythms and channels required for the diagnosis of AD. The proposed framework was tested on a real-time AD EEG dataset. The results reveal that together, the gamma and beta rhythms in the channels, Cz, F4, P4, T6, Pz were the most reliable biomarker for identifying AD and the proposed LSTM based model yielded the best performance. Additionally, another mild cognitive impairment (MCI) EEG dataset was used to test the proposed approach, and the results were excellent (accuracy >99%). The proposed framework will be useful for creating a CAD system to perform automatic AD diagnosis.

Index Terms—Alzheimer's disease (AD), electroencephalography (EEG), biomarkers of EEG, long short-term memory (LSTM), mild cognitive impairment (MCI), feature extraction, classification.

I. INTRODUCTION

ALZHEIMER'S disease (AD) is a neurodegenerative brain disorder that damages the brain's neurons and their connections, resulting in memory loss, cognitive impairment, intellectual deficits, and behavior disorders. AD is a highly disruptive

Manuscript received 8 June 2023; revised 22 August 2023; accepted 24 September 2023. (*Corresponding author: Siuly Siuly*.)

Siuly Siuly is with the Institute for Sustainable Industries & Liveable Cities, Victoria University, Melbourne, VIC 3011, Australia, and also with the Centre for Health Research, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: siuly.siuly@vu.edu.au).

Ömer Faruk Alçin is with the Department of Electrical-Electronics Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, Malatya, Türkiye (e-mail: omer.alcin@ozal.edu.tr).

Hua Wang is with the Institute for Sustainable Industries & Liveable Cities, Victoria University, Melbourne, VIC 3011, Australia (e-mail: hua.wang@vu.edu.au).

Yan Li is with the School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: yan.li@usq.edu.au).

Peng Wen is with the School of Engineering, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: paul.wen@usq.edu.au).

Digital Object Identifier 10.1109/TETCI.2024.3353610

disorder that eventually impedes on a person's ability to carry out daily activities and ultimately results in their death. As the most common form of dementia, accounting for up to 80% of dementia cases [1], [2], AD is the main cause of disability among older people [3]. In Australia, dementia, including AD, is the second leading cause of death [4], and in the USA it is ranked as the third leading cause of death for older people, just behind heart disease and cancer [5]. Globally, there are currently around 40 million people suffering from AD, and this number is predicted to increase to over 130 million by 2030 [6], [7]. According to the World Health Organization [8], in 2019, the total global social cost of dementia was estimated to be US\$1.3 trillion, and this is expected to exceed US\$2.8 trillion by 2030. AD is therefore a challenge for public health and health care systems worldwide, and its economic impact represents a great burden on society.

As there is currently no permanent cure for AD, early detection is the only way to improve the quality of patients' lives, help their family and caregivers ensure appropriate treatments and care is provided, and ensure families are able to access timely information as well as make plans for the future. It is therefore of great importance that an early detection system for the automatic diagnosis of AD is developed. Such a system will help to ensure sufferers of AD can access better health care, and may reduce the impacts of the disease on society and the economy.

The diagnosis of AD can be performed using a variety of procedures, such as lab tests, medical history checks, mental status examination, and, more recently, brain imaging tools like computed tomography (CT) and functional magnetic resonance imaging (fMRI). All these tests are expensive, time consuming, and require experienced clinicians. To avoid the limitations associated with time, cost, and clinician experience, electroencephalography (EEG) technique has recently emerged as a potential tool for AD diagnosis. The technique is simple, non-invasive, low-cost, portable, and has high temporal resolution to recognize inherent mechanisms for the diagnosis of AD compared to other existing techniques [9]. An EEG records the electrical activities generated by brain, and the recorded signals reflect the physiological states of the patient's physical and mental health conditions. EEG signals can also provide valuable information about changes in electrophysiological brain dynamics due to AD.

In recent years, there have been numerous studies focusing on the detection of AD utilizing EEG signal data. Section II presents a concise overview of the recent research conducted on EEG based AD detection [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

From the literature it is evident that most of the methods that have been used to detect AD have not been able to find an adequate balance between accuracy and efficiency. In many of these studies, the same types of features are used in the detection of AD, such as coherence, power modulation spectrogram, complexity features, and FC measures with machine learning techniques. The sizes of the datasets were also small. It is therefore argued that the performances of these methods were similar and not very promising. In addition, in the existing literature there are no studies that investigate the combined impact of EEG rhythms and channels as a potential physiological biomarker for identifying AD.

To address the current deficits identified in research investigating AD detection, this study develops a new framework involving long short-term memory (LSTM) for automatic identification of three types of EEG signals: mild AD, moderate AD, and normal control with improved performance. The framework proposed in this study can be used to identify the ideal combination of rhythms and channels as potential biomarkers for the detection of AD, and can produce more accurate results compared to other methods of AD detection. The motivation for using LSTM in this study is its unique ability to capture sequential patterns and long-term dependencies in sequential data (e.g., EEG data). By leveraging the memory cells and gating mechanisms within LSTM, the model can selectively retain important information from the past observations in the EEG time series data. This allows for the detection of subtle temporal patterns and dependencies that may be indicative of AD-related changes in brain activity. The ability of LSTM to handle long-term dependencies makes it well-suited for capturing the complex dynamics and progression of AD. Furthermore, the combination of optimal EEG rhythms and channels within the LSTM-based model can provide a comprehensive representation of brain activity patterns associated with AD. This makes the proposed approach superior for identifying early stages of AD such as mild AD and moderate AD from EEG signals. To the best of our knowledge, this is the first work to explore the EEG rhythm and channels together using the LSTM network model.

A number of steps were undertaken to develop the framework proposed in this study. First, artifacts from EEG were automatically eliminated by employing a ‘wavelet independent component analysis’ filter. Second, each channel EEG data were decomposed into five rhythms (delta, theta, alpha, beta, gamma) utilizing the ‘wavelet transform (WT)’ technique, and the significant characteristics (in sequential form) were computed from each rhythm for evaluation. Third, a LSTM network model was designed where the obtained sequential characteristics were used as input for identifying different stages of AD (mild AD, moderate AD). The proposed LSTM-based model was then used to choose the optimal combination of EEG rhythms and channels. To check further performance of the biomarkers, the sequential characteristics set obtained in step two were also used

as an input to the multi class support vector machine (multi-class SVM). At the end, the performance of the LSTM model was evaluated by comparing its performance to the performance of the multi-SVM. It was found that the LSTM-based deep learning model performed better compared to the multi-class SVM method.

This study makes several key contributions, including:

- Designing an innovative LSTM-based deep learning framework that utilizes EEG biomarkers from rhythms and channels to automatically identify individuals with different stages of AD (e.g., mild AD and moderate AD).
- Discovering the optimal EEG rhythms and channels that provide crucial information for recognizing different stages of AD.
- Validating the proposed framework by comparing its performance with multi-class SVM.
- Demonstrating improved accuracy levels compared to existing methods of AD detection.

The remainder of this paper is organized as follows. Section II presents a concise literature survey of the current methods used for detecting AD based on EEG signals. The analyzed data and proposed methodology are described in Section III. Section IV presents the information about the experimental set up, the experimental results and their corresponding discussions. Finally, the conclusion and future work is provided in Section V.

II. LITERATURE SURVEY

This section presents an overview of recent research involving the application of LSTM-based RNN models for AD detection using EEG data. This literature review directly aligns with the goal of this study, which helps to identify a well-defined research gap.

In 2023, Ravikanti and Saravanan [17] introduced an Optimized Transformer-based Attention Long Short Term Memory (OTA-LSTM) for AD detection using EEG signals. In their proposed approach, they initially acquired the necessary AD EEG data from an online source. Afterwards, a 3-level ‘Lifting Wavelet Transform (LWT)’ decomposition was applied to these signals to break them down into multiple wavelets. To capture both spatial features and temporal information from the decomposed signal, the researchers employed a combination of RNN and Multi-scale dilated Convolutional Neural Network (CNN). To determine optimal weights, they developed a novel optimization algorithm called the ‘Wild Geese Lemurs Optimizer (EWGLO)’. In the final detection stage, the OTA-LSTM model utilized weighted stacked features, achieving higher accuracy than CNN, RNN, SVM, and A-LSTM (by 4%, 2.6%, 2.5%, and 0.2% respectively). However, the model’s limitations included a restricted subject count for AD detection.

Ho et al. in 2022 [18] developed a deep learning model called DeepADNet for detecting AD using multi-channel EEG signals. They initially extracted event-related spectral perturbation (ERSP) features, which measured average dynamic amplitude changes in three primary EEG frequency bands concerning specific experimental events. Subsequently, they constructed a hybrid deep CNN-LSTM model to leverage ERSP patterns in the

time-frequency domain and generate discriminative features for classifying healthy controls and two AD subject groups. They gathered EEG data from two cognitive ability tests: Oddball (63 subjects, including 23 HC, 17 Pre-symptomatic AD (aAD), and 23 Prodromal AD (pAD) subjects) and N-back (36 subjects, comprising 13 HC, 11 aAD, and 12 pAD subjects). The results indicated significant ERSP pattern differences among the three subject groups across the two experimental stages. The classification outcomes highlighted that the CNN-LSTM model achieved the highest accuracy for the Oddball ($71.95\% \pm 0.019$ and $75.95\% \pm 0.017$) and N-back ($69.40\% \pm 0.003$ and $73.70\% \pm 0.010$). However, the performance of the proposed CNN-LSTM method was relatively low and did not meet practical implementation standards.

In 2022, Alessandrini et al. [19] devised a methodology utilizing principal component analysis (PCA) in conjunction with an LSTM-RNN network to identify AD from EEG signal data. In their research, the authors introduced a robust PCA (RPCA) algorithm to eliminate outlier components from the signals. The primary objective of this study was to validate the effectiveness of applying RPCA to corrupted EEG signals, which aimed to cleanse the signals and subsequently extract features through PCA. The final step involved crafting an LSTM-RNN network model integrated with RPCA for automated AD detection. The investigation employed a dataset from EEG recordings involving 35 hospitalized subjects, including 20 with AD and 15 healthy individuals. The results demonstrated that even when encountering data corruption leading to erasures of up to 20%, RPCA managed to enhance detection accuracy by approximately 5% compared to the baseline PCA. However, a limitation of the study was the absence of comprehensive experimental evaluation to establish parameters for the proposed LSTM model, such as epochs, optimizers, batch size, and hidden neurons.

Gkenios et al. in 2022, [20] designed three distinct Long Short-Term Memory (LSTM) models designed for categorizing subjects into three classes: AD, Mild Cognitive Impairment (MCI), and Healthy Control (HC) groups, using EEG data. The first model, named 'LSTM,' comprises a single LSTM layer with 512 units. The second model, 'Conv-LSTM,' is similar to the first but integrates a 1D Convolutional layer with 512 filters, a kernel size of 3×3 , stride of 1, batch normalization, and ReLU activation before the LSTM layer. The third model, 'Conv-BLSTM,' is similar to the second but incorporates bidirectional LSTM. In their study, feature extraction involved applying Fast Fourier Transform (FFT) on each segment. The method was evaluated on an EEG dataset containing 54 subjects: 18 HC, 18 with MCI, and 18 with AD. Among the three models, the second model (Conv-LSTM) exhibited high accuracy when segment-based data were considered, but notably lower accuracy (67%) when the data were organized by subject.

In 2023, Alvi et al. [21] proposed a deep learning framework based on the Long Short-Term Memory (LSTM) model to effectively distinguish individuals with MCI (an early stage of AD), from HC subjects. The proposed framework comprises four main phases: denoising, segmentation, downsampling, uncovering latent features using the LSTM model, and identifying MCI patients via a sigmoid classifier. The study employed an

EEG dataset from 27 participants (16 HC and 11 MCI subjects), collected from Sina and Nour Hospitals in Isfahan, Iran. The complexity of the proposed framework poses challenges for establishing a stable technology for identifying MCI/early stages of AD using EEG signals.

In 2020, Sridhar and Manian [22] introduced a deep Bidirectional LSTM (BLSTM) network model. This model aimed not only to differentiate between individuals with MCI and HC, but also to evaluate cognitive functions and establish their correlation with brain signal features. The signal data matrix underwent initial noise reduction through a Butterworth bandpass filter, targeting frequencies between 3 Hz and 30 Hz. Afterward, dimensionality reduction was performed via PCA. Three principal components derived from PCA constituted the feature matrix, which subsequently served as training input for the BLSTM network. For this investigation, a dataset encompassing 35 subjects (18 female and 17 male) was utilized. Out of these, 28 were healthy subjects, while the remaining seven exhibited mild memory dysfunction indicative of MCI. The proposed model achieved accuracy rates ranging from 69.53% to 91.93% for MCI vs. HC subjects across various age groups. However, these accuracies were considered insufficient for real-time application scenarios.

Petrosian et al. [23] investigated the utilization of EEG data analysis to detect AD, employing a model that combined wavelet and RNN approaches. This method was evaluated on a dataset comprising EEG data from 10 early onset AD patients and 10 control subjects. The optimal results were obtained by using a three-layer RNN model on the fourth-level high wavelet subbands in the left frontal channel. The performance substantially surpassed random guessing, achieving around 80% sensitivity. However, the dataset used for evaluation was limited in size, which hampers drawing broad conclusions. Additional information is required to ascertain the accuracy of this approach across various disease stages and considering treatment effects.

Based on the exploration of the recent literature above, it is evident that none of the research studies employing LSTM models to discover the optimal EEG rhythms and channels required for effective early-stage AD diagnosis. These biomarkers play a pivotal role in comprehending the important brain activity patterns associated with AD. Furthermore, a majority of the LSTM models lacked to select crucial parameters like epochs, optimizers, batch sizes, and hidden neurons through extensive experimental evaluation, resulting in unsatisfactory performance. To tackle these challenges, this study intends to create an innovative LSTM based methodology integrating essential EEG biomarkers to enhance performance of early detection of AD.

III. METHODOLOGY FRAMEWORK

The key objective of this study is to present a new LSTM-based framework for the automatic detection of different stages of AD patients using optimal EEG rhythms and channels. The proposed framework consists of four steps: (i) automatic artifact removal, (ii) rhythm decomposition and characteristics abstraction, (iii) design of the LSTM network model for deep feature

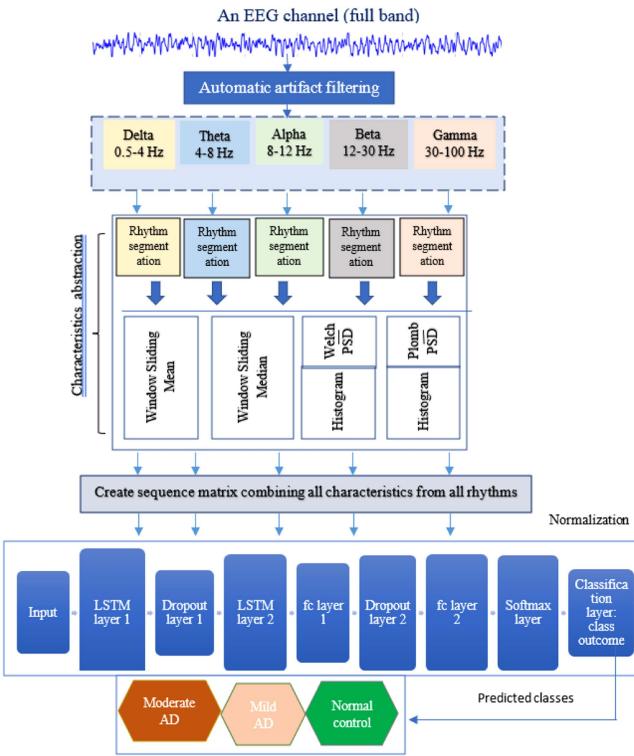


Fig. 1. Overview of the proposed LSTM based framework for identifying different stages of AD from EEG.

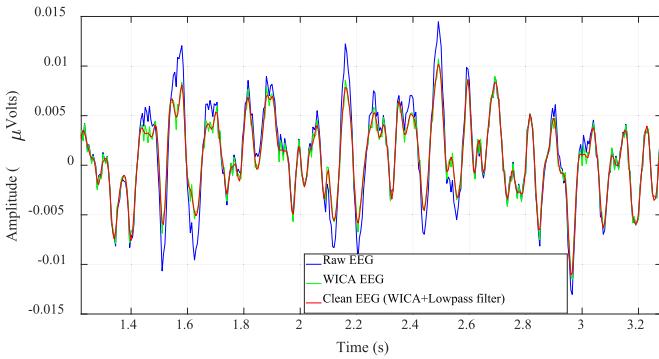


Fig. 2. Illustration of a raw EEG signal and artifact free EEG signal.

extraction and classification, (iv) performance evaluation and decision making. Fig. 1 presents an overview of the proposed framework. Further description of the data and the steps used in the framework are provided in the following sections.

A. Data Acquisition

This study used an EEG dataset from AD patients, which was collected from The Behavioral and Cognitive Neurology Unit of the Department of Neurology and The Reference Center for Cognitive Disorders at the Hospital das Clínicas in São Paulo, Brazil. The dataset was compiled from the data of 86 participants. Of these participants, 31 had mild AD (AD1), 20 had moderate AD (AD2), and 35 were normal control (NC)

TABLE I
DEMOGRAPHICS INFORMATION OF THE PARTICIPANTS

	NC	AD1	AD2
Subject (male/female)	35 (16/19)	31 (12/19)	20(6/14)
Age (years)	66.89±8.18	75.23±5.55	73.77±10.16
Education (years)	8.77±5.28	4.81±2.89	4.73±3.87
MMSE score	28.00±2.20	19.48±3.16	14.18±3.69
CDR score	0	1	2

NC=Normal control; AD1=Mild AD; AD2=Moderate AD; MMSE= Mini-Mental State Examination; CDR=Clinical Dementia Rating;
*average±standard deviation.

subjects. AD diagnosis was performed by experienced neurologists based on the Brazilian version of the Mini-Mental State Examination (MMSE) [24] and the Clinical Dementia Rating (CDR) scale [25]. The recording of EEG signals was acquired from 20 channels with the participants awake, relaxed, and with their eyes closed for at least 8 min (12-bit resolution and 200 Hz sampling frequency). The electrodes were placed according to the international 10-20 montage system. A detailed description of this dataset is available in [14], [26]. Table I presents the demographic characteristics of the participants.

B. Automatic Artifact Removal Using Wavelet Independent Component Analysis Filter

Artifacts are unwanted signals that are generated from environmental noise, experimental error, and physiological issues (e.g., eye blinks, eye movements, muscle movement) during EEG recording. As these artifacts contaminate the quality of the original EEG signals and may lead to incorrect conclusions in the analysis, it is essential that unwanted signals are cleaned from EEG data. Thus, in the first step of our proposed framework, we aimed to eliminate artifacts from the original EEG signals using an automatic artifact removal technique. In this study, we employed a crossbreed filter, called ‘Wavelet Independent Component Analysis (WICA)’ for the elimination of unwanted signals from the EEG. Afterwards, we filtered the EEG recordings with a lowpass filter at 100 Hz to properly clean the EEG data in preparation for the next step in the framework. The reason for using the WICA filter in this study was the capability of the method for designing the data into a new space where the redundancy was higher and the characteristics of the artifacts in the frequency domain were fully exploited. This method also has ability to better preserve EEG properties, reducing information loss and signal distortion. Previous studies [14], [27], [28] have applied a similar type of technique for cleaning artifacts.

The WICA method involves the following steps:

- 1) Initially, apply conventional independent component analysis (ICA) decomposition to the raw EEG data which yields the mixing matrix (M) and N independent components.
- 2) Perform a wavelet transform on the independent components to produce the wavelet coefficients.
- 3) Apply a thresholding technique to the wavelet coefficients, identifying and selecting those coefficients that exceed the threshold.

- 4) Perform an inverse wavelet transform on the thresholded coefficients, reconstructing the components that solely consist of neural sources $\{n_i(t)\}$.
- 5) Create the wICA-corrected EEG by combining the reconstructed components: $\hat{X}(t) = M[n_1(t), n_2(t), \dots, n_N(t)]$.

In this study, we chose the ‘Coiflets’ wavelet family and the 5 levels of decomposition, such as ‘coif5’, for the WICA approach. Fig. 2 shows an example of the pattern of a raw EEG signal compared to the pattern of that signal after artifact removal artifact. The signal presented in Fig. 2 is of the F7 channel signal from a moderate AD patient. After the WICA and lowpass filtering, we acquired clean EEG data that was used for the next step.

C. Rhythm Decompositions and Characteristics Abstraction

Physiological rhythms of the EEG signals play an important role in detecting AD patients. This is because physiological rhythms hold useful insights about brain functionality and synchronization. Each rhythm conveys distinctive physiological information on brain functional state. To extract useful information for AD diagnosis, we used wavelet decomposition to divide each EEG channel signal into five rhythms, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz). we chose to adopt 5 levels of decomposition using the ‘Daubechies 5’ wavelet (e.g., db5) because db5 is widely recognized as one of the most successful wavelet types to capture the essential frequency bands inherent in EEG signals (e.g., delta, theta, alpha, beta, and gamma) [29]. Each level of decomposition corresponds to a specific frequency range. An example of the decomposition for an EEG channel and the separated rhythms are exhibited in Fig. 3. We investigated which rhythm/rhythms captured the most discriminative information for efficient identification of AD.

To obtain representative information from different rhythms for AD, we computed four important characteristics, including ‘mean’, ‘median’, ‘Lomb-Scargle periodogram Power Spectral Density (LPSD)’, and ‘Welch’s power spectral density (WPSD)’ from each rhythm. The reasons of considering these characteristics are that the mean and median were particularly important for describing the distribution of data in each rhythm, and LPSD and WPSD were important for presenting signal power distribution over frequency to preserve the frequency resolution, as well as to show power variations (energy) as a function of frequency. Both the LPSD and the WPSD deliver an excellent estimate of the spectral power at the cost of low computational complexity.

We wanted to design a Long Short-Term Memory (LSTM) deep learning model for evaluating the obtained representative characteristics from different rhythms as input. According to the LSTM model requirement, the input vector must be a sequence. To improve the temporal resolution of EEG and obtain better performance, we segmented each rhythm of each channel into 16 non-overlapping windows (window length was chosen with 100 sample empirically) and then calculated the ‘mean’/‘median’ for each of the 16 windows. Fig. 4 presents an example how we calculated mean and median from each window of a rhythm to

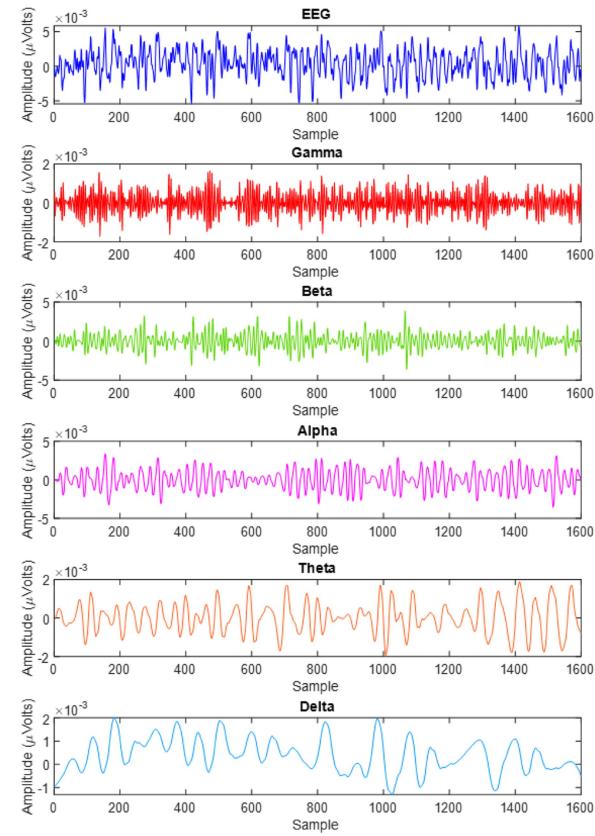


Fig. 3. Illustration of patterns of an EEG channel signal and the separated five rhythms.

create a sequence of size 1×16 . In Fig. 4, V_i denotes statistical parameter (mean/median) of i th window. This means that for a rhythm, we obtained a sequence of 1×16 size a median characteristic. Thus, for mean and median characteristics, we obtained a sequence of 2×16 for a rhythm. Using LPSD and WPSD, we computed spectral power from each rhythm and considered the histogram bins as 16 to form a 1×16 sequence to adjust for the size of the sequence in line with other characteristics. Fig. 5 presents an example of how LPSD or WPSD is calculated from each rhythm. Thus, for LPSD and WPSD characteristics, we obtained a sequence of 2×16 size for a rhythm. This means that for four characteristics (mean, median, LPSD, and WPSD), we obtained a sequence of 4×16 for a rhythm.

To illustrate clear picture of the process used in combining characteristics, an example is provided in Fig. 6. This shows how the significant information is generated from each channel for the LSTM model. As an example, in Fig. 6, we considered two rhythms (gamma and beta) from a channel (Cz) and obtained a sequence of 4×16 (number of characteristics \times number of windows) for each rhythm of the Cz channel. For gamma and beta rhythms in the Cz channel, we obtained a sequence of 8×16 . Thus, for five rhythms (delta, theta, alpha, beta, gamma) of a channel (e.g., Cz), we can obtain a sequence of 20×16 size. As mentioned in section A, the dataset used in this study was obtained from EEG recording signals from 20 channels for 86 participants. For each participant, we obtained a sequence

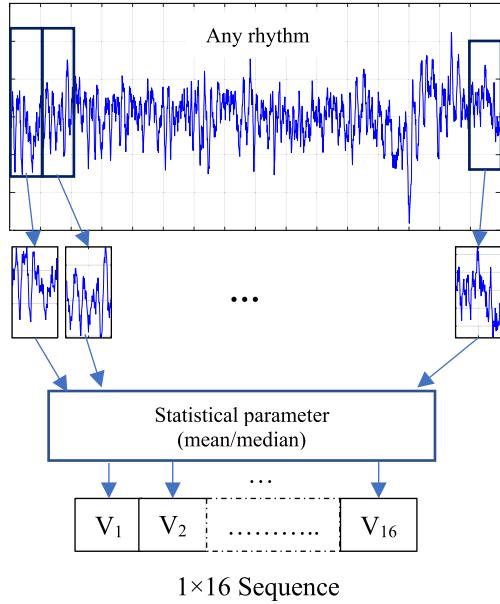


Fig. 4. Illustration of creating sequence vector from windows of a rhythm considering mean/median parameter.

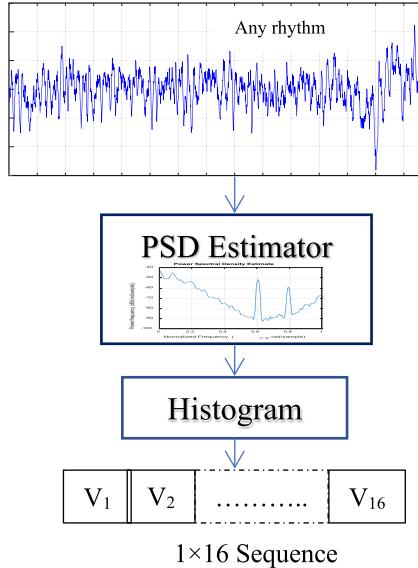


Fig. 5. Illustration of creating sequence vector from histogram bins of a rhythm considering LPSD/WPSD.

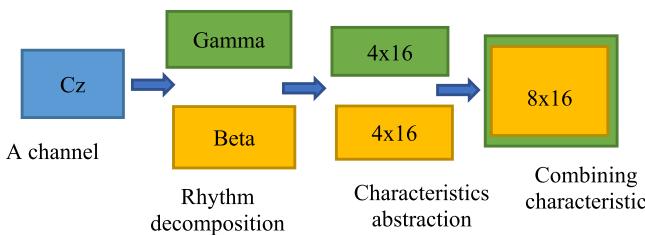


Fig. 6. Example of generating process of significant characteristics from a channel.

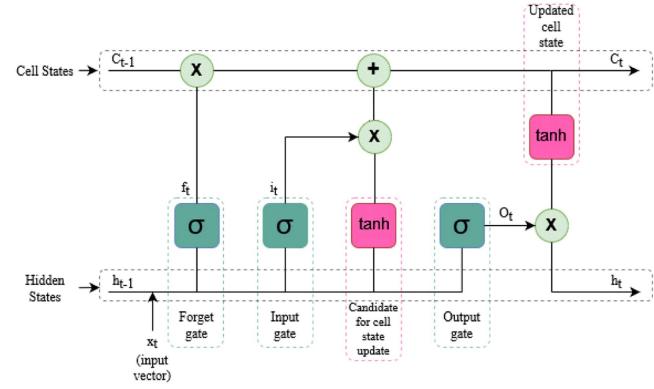


Fig. 7. Architecture of a LSTM cell unit.

of (400×16) involving five rhythms and 20 channels. Using the data for the 86 participants, we created a complete sequence matrix that represented all 86 participants, and then performed normalization (with zero mean and unit variance (z-score)). The sequence matrix was used as input to the LSTM model in the next step.

D. Design a LSTM Architecture Based Model for Detection of Different Stages of AD

Long Short-Term Memory Network (LSTM) is an advanced recurrent neural network (RNN) that is used for modelling sequence data (such as time series data (e.g., EEG data)) for building predictive models [30], [31]. The use of the LSTM allows to connect previous information to the present task to better capture long-term dependencies. While traditional RNNs can learn short-term dependencies, they often have difficulties in learning long-term dynamics due to the vanishing and exploding gradient problems. LSTM is a type of RNN that addresses the vanishing and exploding gradient problems by learning both long- and short-term dependencies [30], [31], [32], [33], [34], [35].

A LSTM network model consists of a ‘memory cell (cell state)’, ‘forget gate’, ‘input gate’, and ‘output gate’. The memory cell of the LSTM layer stores or remembers values (states) for either long or short time periods [31], [32]. Fig. 7 shows an architecture of a LSTM model how the components of an LSTM block work in a cell unit. In a cell of the LSTM network, the forget gate decides whether the information is thrown from the previous time step shown in (1). In (1), if f_t is 0 then the network will forget everything, and if the value of f_t is 1 then it will forget nothing. The input gate controls the amount of the new information (value) kept in the cell, as shown in (2). Here the new information in the cell state is a function of the hidden state at the previous time step and the current time step, as shown in (3). The output gate controls the amount of the value in the cell, to be used for computing the output as shown in (4). Each gate contains a fully connected layer and an activation function.

$$\text{Forget gate}, f_t = \sigma(w_f [x_t, h_{t-1}] + b_f) \quad (1)$$

$$\text{Input gate}, i_t = \sigma(w_i [x_t, h_{t-1}] + b_i) \quad (2)$$

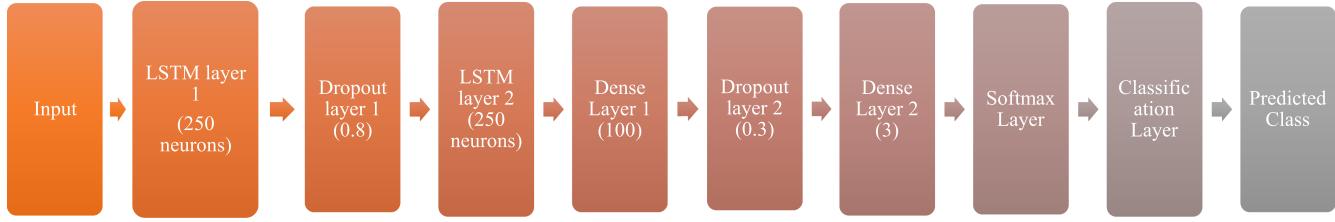


Fig. 8. Proposed LSTM network design model for identification of different stages of AD.

$$\text{New information, } N_t = \tanh(w[x_t, h_{t-1}] + b) \quad (3)$$

$$\text{Output gate, } o_t = \sigma(w_o[x_t, h_{t-1}] + b_o) \quad (4)$$

$$\text{Updating cell state, } c_t = f_t \times c_{t-1} + i_t \times N_t \quad (5)$$

$$\text{Current hidden state, } h_t = o_t \times \tanh(c_t) \quad (6)$$

$$\text{Current output, } y_t = \text{Softmax}(h_t) \quad (7)$$

Where, $x_t = (x_1, x_2, \dots, x_T)$ is an input sequence and $y_t = (y_1, y_2, \dots, y_T)$ is an outcome sequence and $t = 1$ to T . $\sigma(x) = \frac{1}{1+e^{-x}}$; $\tanh = \frac{2}{1+e^{-2x}} - 1$; h_t is the hidden state at time step t , previously hidden state h_{t-1} , candidate value of a new cell c_t , w_f, w_i, w, w_o are the weights, and b_f, b_i, b, b_o are the biases.

In this study we built a prediction model based on LSTM design to identify different stages of AD. Equation (7) was used to estimate the progression of AD. In the proposed model, we used 20×16 sequence for each channel as an input to the network. We have already discussed the procedure related to how we obtained this input sequence set (Section C). Our proposed LSTM network architecture consisted of two LSTM layers (LSTM layer 1 and LSTM layer 2) with 250 neurons each. The model also included two dropout layers, with a dropout rate of 0.8 for dropout layer 1 and 0.3 for dropout layer 2. Additionally, there were two dense layers with 100 neurons for dense layer 1 and 3 neurons for dense layer 2. The output was obtained through a softmax layer and a classification layer for a three-class problem (e.g., AD1, AD2, NC). The design of the proposed LSTM model is shown in Fig. 8.

The performance of the LSTM network model depends on a number of hyper-parameters, including learning rate, number of LSTM hidden neurons, batch size, number of training epochs, and dropout rates. In this study, we performed a systematic search and investigation that considered several combinations of the parameters to determine the appropriate value of the hyperparameters. After empirical evaluation, we trained our proposed LSTM model through an adam optimizer with 0.01 learning rate, and 128 (minibatch size) sequence per iteration at 100 epochs with hidden 250 neurons. To avoid the overfitting of the LSTM network model, we adopted two dropout layers where first dropout rate was 0.8 and second one was 0.3.

In order to discover the most responsible rhythms and channels for the detection of the different stages of AD, this study introduces a strategy that included the following steps: (1) calculate overall accuracy for each of five rhythms in each channel; (2) identify the optimal rhythm/rhythms (selected rhythm/rhythms)

by choosing the rhythm/rhythms for which the overall accuracy was greater than 80% for most of the channels; (3) select the channels that produce the highest performances with any separate rhythm or all rhythms or selected rhythms.

Following this strategy, we performed a wide range of experiments (the results of these are reported in Section III-B). Finally, we found that two rhythms, “gamma and beta”, with the five channels, “Cz, F4, P4, T6, Pz”, jointly produced a higher performance compared to other rhythms. Thus, the results revealed that “gamma and beta” and “Cz, F4, P4, T6, Pz” channels possessed the most accountable biomarkers for identifying AD using EEG.

To further assess the performance of the obtained sequence feature set, we applied the same feature set to a multi-class SVM (one vs one) model [35]. The motivation of considering the multi-class SVM was that this method is very popular and reliable in the machine learning community to deal with high dimensional data, and it has the capability to produce good performance. The multi-class SVM model was implemented in MATLAB and a linear kernel was used as the kernel function (chosen after testing all kernels), and the box constraint parameter selected was 1.5 (after empirical evaluation).

To rigorously evaluate the performance of the proposed models, most of the standard measurements such as accuracy, sensitivity, specificity, precision, F1 score, Matthews correlation coefficient (MCC), and kappa coefficient and operating characteristic curve (ROC) were considered in this study. Descriptions of these performance measurements are available in [36], [37].

IV. EXPERIMENTS AND RESULTS

This section provides a brief description of the experimental set up and implementation process, and presents the obtained results with the corresponding discussion.

A. Experimental Set Up

To verify the validity of the proposed framework, this study used a publicly available EEG Alzheimer's disease dataset. The dataset contained 86 participants, where 31 subjects had mild AD (AD1), 20 had moderate (AD2) and 35 subjects were healthy control (NC). Each participant's data included 8-second epochs of 20-channel EEG, recorded at 200 Hz sampling rate (1600 samples per epoch). For the 86 participants, a total of 1514, 930, and 1426 epochs were obtained for AD1, AD2 and NC subjects, respectively. Thus, the total dimension matrix for the AD1, AD2, and NC EEG epochs was obtained as $1514 \times 20 \times 1600$,

TABLE II
OVERALL ACCURACY (IN PERCENTAGE) FOR EACH CHANNEL BASED ON RHYTHMS FOR THE PROPOSED LSTM-BASED MODEL FOR AD EEG DATASET

Channel	Gamma	Beta	Delta	Alpha	Theta	All Rhythm	Selected Rhythm (gamma & beta)	p values for selected rhythm
F7	82.83	83.56	71.04	55.74	40.43	89.14	91.42	0.0000*
T3	85.94	81.28	70.42	53.57	42.71	87.28	87.28	0.0000*
T5	87.28	83.25	62.98	55.64	43.54	85.73	89.04	0.0000*
Fp1	84.59	81.28	71.98	55.64	40.95	86.56	90.49	0.0000*
F3	86.66	85.73	65.77	39.09	41.99	90.59	36.93	0.10150
C3	84.90	85.94	67.53	44.36	42.81	89.87	92.45	0.0000*
P3	83.66	84.07	68.15	52.74	41.68	87.28	92.04	0.0000*
O1	83.25	86.76	62.36	58.32	42.61	88.31	92.86	0.0000*
F8	86.04	86.35	74.77	51.40	42.19	89.66	90.49	0.0000*
T4	84.59	84.07	68.25	57.19	42.71	89.35	91.62	0.0000*
T6	87.49	85.63	68.25	52.74	44.98	85.73	93.49	0.0000*
Fp2	84.69	84.59	71.35	53.05	39.71	89.35	91.31	0.0000*
F4	86.56	82.01	71.25	52.64	39.40	87.49	93.69	0.0000*
C4	84.28	80.46	63.29	55.84	42.30	86.87	89.87	0.0000*
P4	86.76	82.42	62.87	54.81	41.99	85.83	93.59	0.0000*
O2	86.87	83.45	63.50	59.98	44.88	91.52	92.76	0.0000*
Fz	39.09	39.09	58.95	51.09	39.81	83.45	91.62	0.0000*
Cz	83.87	39.09	65.56	57.81	43.43	88.73	93.90	0.0000*
Pz	86.25	85.73	66.08	53.88	41.47	85.21	93.38	0.0000*
Oz	84.18	86.04	60.29	53.57	38.68	88.42	91.62	0.0000*

$930 \times 20 \times 1600$, $1426 \times 20 \times 1600$ (epochs x number of channels x number of samples), respectively. The whole dataset size, combining the three groups together was $3870 \times 20 \times 1600$. In this study we divided the whole dataset into two parts: 75% (2903 epoch) for training purpose and 25% (967 epoch) for testing purpose. In all experiments, the training data set was used to train the proposed scheme and the testing dataset was used to test the performance of the proposed model. All the experiments in this study were carried out in a MATLAB environment, on a computer with a 2.60 GHz Six-Core CPU, 32 GB RAM, and NVIDIA RTX 2070 GPU with 8 GB of memory.

B. Results

This study presents a LSTM-based network model that utilizes EEG signal data, including EEG rhythms and characteristics, to differentiate different stage of AD: mild AD (AD1), moderate AD (AD2) from normal control (NC) subjects. The research focuses on identifying the most informative rhythms and channels for effective AD detection. To determine the optimal rhythms within the channels, we calculated the overall accuracy for each channel for each channel using different rhythms, as shown in Table II. In addition, we conducted an analysis considering all rhythms (such as delta, theta, alpha, beta, gamma) together and computed the overall accuracy for each channel. As shown in Table II, gamma and beta individually exhibited superior accuracy (over 80%) compared to other rhythms across nearly all channels. Consequently, we selected gamma and beta as the combined “selected rhythm” and evaluated its performance for each channel. Remarkably, the combined gamma and beta (selected rhythm) yielded the highest accuracies for almost all channels.

To validate the significance of the “selected rhythm,” we performed a one-way analysis of variance (ANOVA) to determine if the selected rhythm for each individual channel has statistical significance in identifying different stages of AD at a

5% level of significance. The resulting probability (p) obtained from ANOVA is presented in the last column of Table II. Upon examining Table II, it is evident that the p -values for the selected rhythm with each channel are less than 0.05 ($p \leq 0.05$), indicating that the selected rhythm is a significant factor in distinguishing AD from NC participants for each channel, except for the F3 channel. The F3 channel lacks significance for AD detection from EEG, possibly because it is not located in the region of the brain most affected by AD pathology or may not capture the crucial biomarkers associated with AD. Variability in brain anatomy, electrode placement, and dataset characteristics can also influence the significance of a channel.

To clearly visualise the performance among the channels, Fig. 9 displays a transparent picture. Fig. 9(a) presents a divergence of the accuracy values for the different rhythms and for each of the channels, and Fig. 9(b) shows the comparison between all rhythms and selected rhythms for each of the channels. It is apparent from Fig. 9(a) that the accuracy for the gamma and beta bands are significantly higher compared to the low frequency bands such as delta, alpha and theta, and this is the case for all of the channels, except Fz and Cz. On the other hand, the lowest performances observed were for theta rhythm for each of the channels. From Fig. 9(b), it is obvious that the selected rhythm: ‘gamma and beta band’ jointly yields the best accuracy when compared to ‘one rhythms’ and individual rhythms for all of the channels except the F3 channel. The results demonstrate that the top two higher frequency bands, ‘gamma and beta’ together possess significant information related to brain changes associated with AD and are therefore important in the recognition of different stages of AD using EEG.

Fig. 10 displays a visualisation of gamma band features for a mild AD (AD1), moderate AD (AD2), and healthy (HC) participant for each of the channels in the AD EEG dataset. This diagram illustrates an example how a rhythm (frequency band) characteristics can be used to distinguish distinct categories of AD.

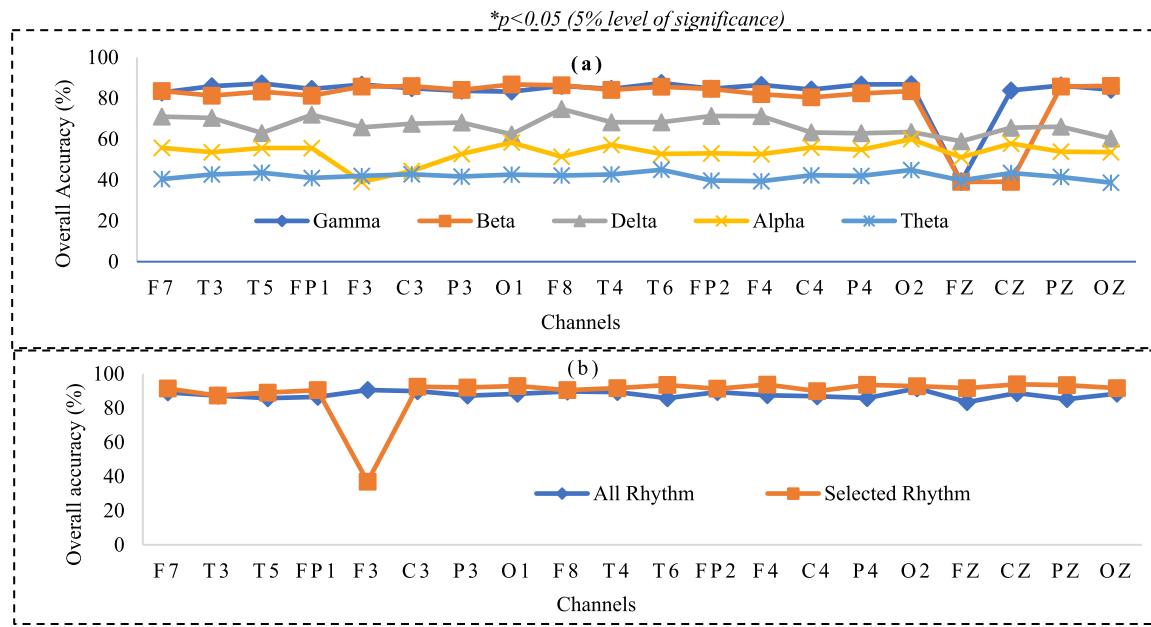


Fig. 9. Performance comparison (a) channel vs different rhythms (b) channel vs all rhythm/selected rhythm.

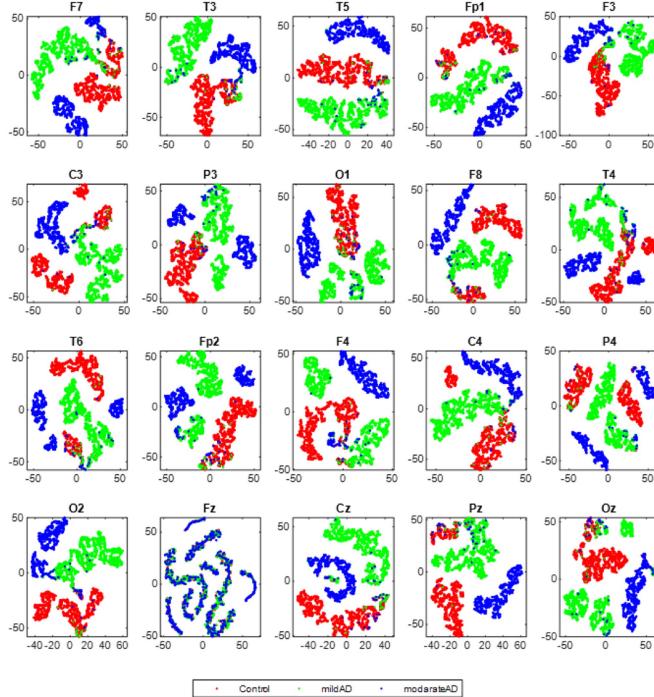


Fig. 10. Exemplary feature visualization for gamma band in each 20 channels of AD EEG dataset.

Based on the Fig. 10, it is evident that T5, Fp1, T3, and Cz channels in the gamma band exhibit prominent discriminating features for identifying AD1, AD2, and HC subjects. In contrast, Fz channel features are indistinguishable, possibly due to the neglect of interactions and connectivity patterns between the prefrontal cortex and other brain regions, which are relevant for AD detection.

To determine the optimal combination of channels and rhythms for AD detection, we conducted a comprehensive set of experiments for the proposed LSTM-based model. The classification performance of the model was evaluated in terms of accuracy, sensitivity, specificity, precision, F1 score, MCC, and kappa coefficient, as summarized in Table III. Table III presents the performances in two ways: (1) considering ‘all (20) channels’ with various combinations of rhythms, and (2) focusing on ‘selected (5) channels’ with various rhythms. For the ‘selected 5 channels,’ we identified the channels from each rhythm that achieved the highest accuracy values. For example, in the ‘gamma’ rhythm, the channels ‘T6, T5, O2, P4, F3’ demonstrated superior performance compared to other channels. Hence, we included the ‘gamma (T6, T5, O2, P4, F3)’ combination in Table III. Throughout this process, we selected 5 channels for each of the rhythms as outlined in Table III.

As can be seen from Table III, in every case, the ‘selected channels’ yield better performance with the different rhythms compared to ‘all channels’. It is also observed that the selected rhythms (gamma and beta together) produce superior performances compared to other rhythms in both cases (all channels and selected channels). As shown in Table III, the highest performances for the proposed LSTM framework were achieved for the selected 5 channels, ‘Cz, F4, P4, T6, Pz’, with the selected rhythms, ‘gamma and beta’, where accuracy, sensitivity, specificity, precision, F1 score, MCC value, and kappa coefficient values are 97.00%, 96.91%, 98.41%, 97.33%, 97.10%, 95.60% and 93.30%, respectively. In both cases (all channels and selected channels), the lowest performances were produced for the theta band with the selected channels, ‘T6, O2, T5, Cz, C3’. This was expected as this low frequency band does not carry out substantial information for identifying AD. The experimental results revealed that ‘T6, O2, T5, Cz, C3’ channels with the gamma and beta band together are highly

TABLE III
CLASSIFICATION RESULTS (IN PERCENTAGE) FOR THE PROPOSED LSTM NETWORK MODEL FOR AD EEG DATA

Channels	Various Rhythms	Accuracy	Sensitivity	Specificity	Precision	F1 score	MCC	Kappa
All Channels (20 channels)	Gamma	94.11	94.51	97.16	93.52	93.90	91.00	86.70
	Beta	95.14	94.64	97.54	94.96	94.80	92.40	89.10
	Delta	90.28	89.14	94.98	90.37	89.60	84.80	78.10
	Alpha	85.42	85.00	92.51	85.35	85.10	77.70	67.20
	Theta	65.36	65.42	82.23	65.25	65.30	47.60	22.10
	All Rhythms	90.28	89.88	94.87	91.08	90.40	85.40	78.10
	Selected Rhythms (Gamma & Beta)	96.38	96.19	98.08	96.75	96.40	94.60	91.90
Selected 5 Channels	Gamma (T6, T5, O2, P4, F3)	96.69	96.69	98.33	96.53	96.60	94.90	92.60
	Beta (O1, F8, Oz, C3, F3)	96.48	96.30	98.18	96.53	96.40	94.60	92.10
	Delta (F8, Fp1, Fp2, F4, F7)	92.24	91.99	96.12	91.75	91.90	88.00	82.50
	Alpha (O2, O1, Cz, T4, C4)	81.90	81.54	90.62	82.36	81.80	72.70	59.30
	Theta (T6, O2, T5, Cz, C3)	57.39	56.16	77.87	57.95	56.70	35.00	4.10
	All Rhythms (T6, F8, O2, F4, P3)	93.28	92.96	96.53	93.32	93.10	89.70	84.90
	Selected Rhythms (Gamma & Beta) (Cz, F4, P4, T6, Pz)	97.00	96.91	98.41	97.33	97.10	95.60	93.30

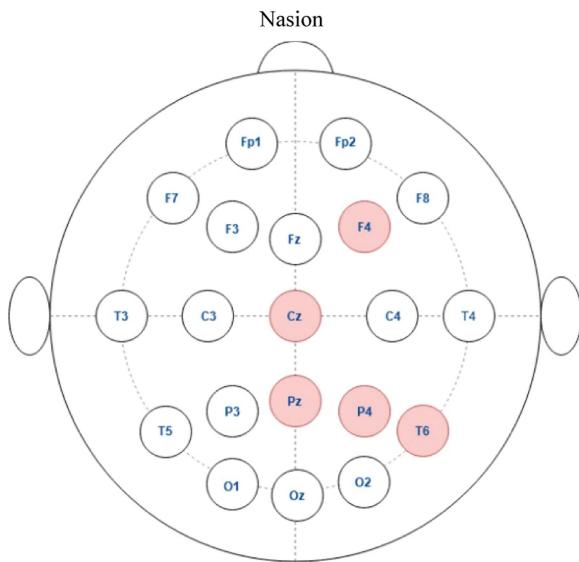


Fig. 11. Responsible channels in the combined gamma and beta rhythm for identifying severity of AD.

responsive to conveying important information for the detection of the different stages of AD. This finding is also consistent with the biological point of view. For example, the gamma band is recorded in the somatosensory cortex and the beta band is observed in the parietal and frontal region of the scalp. These two brain regions are associated with, and memory processing, cognitive generation, social behaviour, language, and AD changes these areas. Fig. 11 illustrates a finding from this study related to the positions of the selected 5 channels in the scalp of the brain. They are highlighted using the colour red.

To demonstrate an exemplary pattern of channel correlation, a correlation matrix is built in Fig. 12 using five channels ('Cz, F4, P4, T6, Pz') which were chosen as the most favourable for AD detection (see Fig. 11). In Fig. 12, a functional linear relationship between two channels is shown by the correlation values (left top in each cell except diagonal cells) (bigger value indicates higher relationship). The correlation between the signals varies from slightly to moderately correlated (-0.09 to 0.77), [43] but none

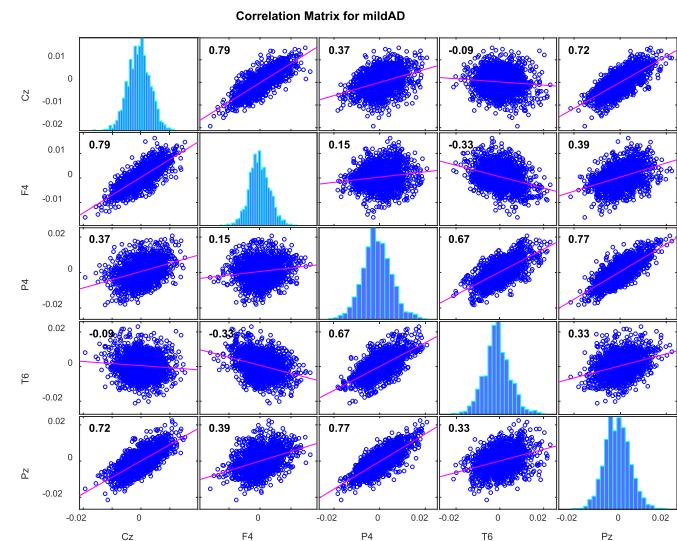


Fig. 12. Correlation matrix for 'Cz, F4, P4, T6, Pz' channels from a mild AD subject.

of them are highly connected, as shown in Fig. 12. If a correlation value between the two signals is greater .85 or less than -.85, then they are strongly correlated [43]. If this occurs, we may not need to include both of the channels/signals for the classification as they will not add any significance in the classification process. Therefore, we can say from the correlation matrix that all the five signals/channels are important to detect AD.

To further verify the effectiveness of the proposed LSTM model, the feature set applied to the LSTM model was then applied to a multi-class support vector machine (multi-class SVM), using the same one vs one approach that was used in the LSTM model for the classification of AD1, AD2 and NC. Fig. 13(a) presents a confusion matrix for the LSTM classifier and Fig. 13(b) for the multi-class SVM classifier. Fig. 13(a) & (b) show that compared to the multi-class SVM classifier, the classified numbers for the LSTM classifier are higher in each of the three classes: AD1, AD2 and NC. Fig. 13(c) provides class-wise accuracy and F1 score for the proposed LSTM based model and

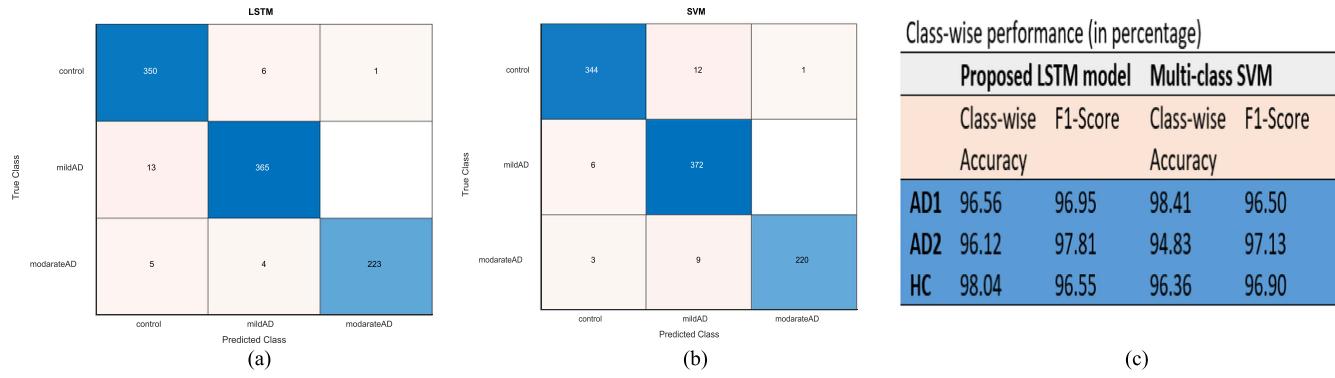


Fig. 13. (a) Confusion matrix for the LSTM classifier. (b) Confusion matrix for the multi-class SVM classifier. (c) Class-wise performances.

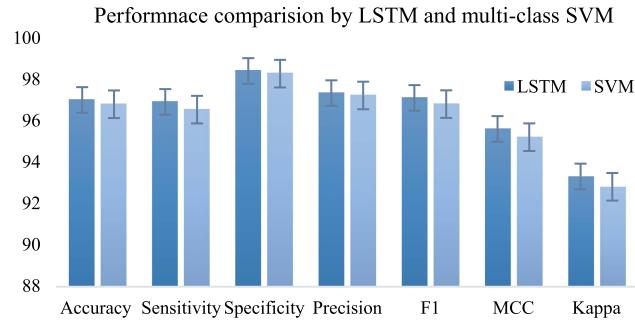


Fig. 14. Performance comparison of the proposed feature set by the LSTM and SVM model.

multi-class SVM method. As can be seen in Fig. 13(c), in most of the cases, the class-wise accuracies and F1-scores are superior for the LSTM model compared to the multi-class SVM method.

To compare the performances between the LSTM model and multi-class SVM model, we calculated the performance parameters for the multi-class SVM model in term of accuracy, sensitivity, specificity, precision, F1 score, MCC, and kappa coefficient and these are presented in Fig. 14. As Fig. 14 shows, in all cases, the LSTM model with the proposed biomarkers produces better performances than the multi-class SVM classifier.

To further assess the performance of the proposed LSTM model by comparing it to the multi-class SVM model, Fig. 15(a) presents a ROC curve for the proposed LSDM model and Fig. 15(b) for the multi-class SVM model. An overall performance of a model is measured through the area value under the ROC curve which lies between 0 and 1 (a larger area value reveals better performance of the classifier). It is apparent from Fig. 15(a) and (b) that the area values of the proposed LSTM model for each of three categories: control, mild AD, and moderate AD, are higher compared to the multi-class SVM model. Thus, the experimental results show that the LSTM-based model that used the selected EEG rhythms (gamma and beta) and channels (T6, O2, T5, Cz, C3) as biomarkers has an excellent ability to efficiently identify AD1, AD2 and NC using EEG data.

In addition, we investigated the performance of the proposed LSTM based scheme using another EEG dataset from mild cognitive impairment (MCI) patients. MCI is an indicator to

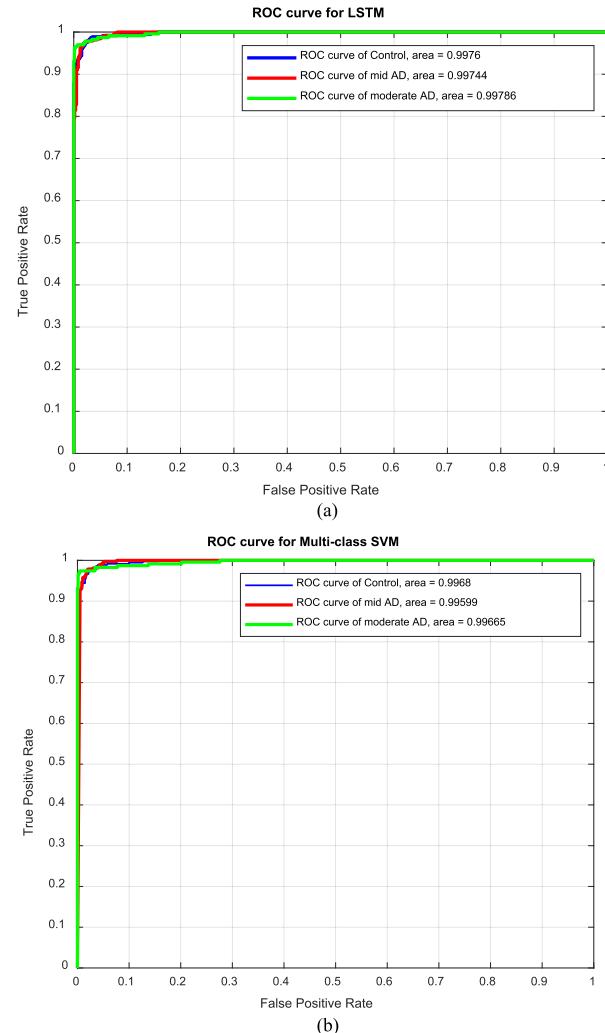


Fig. 15. (a) ROC curve for the proposed LSTM. (b) ROC curve for the multi-class SVM model.

represent the early stage of AD. The data set included EEG recordings from 27 subjects, 16 of whom were cognitively healthy and 11 of whom were MCI patients (aged from 60 to 77 years). The EEG signals were recorded at a 256 Hz sample rate from 19 channels for 30 minutes. Reference [9], [14] provides

TABLE IV
CLASSIFICATION RESULTS (IN PERCENTAGE) FOR THE PROPOSED LSTM BASED NETWORK MODEL FOR MCI EEG DATA

Channels	Various Rhythms	Accuracy	Sensitivity	Specificity	Precision	F1 score	MCC	Kappa
All Channels (19 channels)	Gamma	87.16	83.85	91.30	92.33	87.89	74.69	74.30
	Beta	60.00	76.59	39.26	61.18	68.03	17.12	16.37
	Delta	92.51	93.04	91.85	93.45	93.24	84.84	84.84
	Alpha	61.89	72.00	49.26	63.95	67.74	21.83	21.60
	Theta	55.64	92.30	9.81	56.13	69.80	3.73	2.30
	All Rhythms	97.04	97.48	96.48	97.19	97.34	94.00	94.00
	Selected Rhythms (Gamma & Delta)	98.44	98.52	98.33	98.66	98.59	96.83	96.83
Selected 5 Channels	Gamma (C3, Cz, Fp1, Fz, Pz)	97.45	98.52	96.11	96.94	97.72	94.84	94.82
	Beta (O1, Pz, Cz, T6, F8)	83.46	79.11	88.89	89.90	84.16	67.60	67.00
	Delta (Cz, Fp1, F4, T5, F8)	95.47	97.19	93.33	94.80	95.98	90.84	90.80
	Alpha (T5, O2, T6, Cz, Fp2, C4)	76.79	79.26	73.70	79.03	79.14	52.98	52.98
	Theta (Fp1, O2, T6, Fp2, C4)	71.93	83.70	57.22	70.98	76.82	42.78	41.86
	All Rhythms (C3, Fz, O1, T5, T4)	98.35	98.37	98.33	98.66	98.52	96.67	96.67
	Selected Rhythms (Gamma & Delta) (C3, Fp1, Fz, F8, Cz)	99.09	98.67	99.63	99.70	99.18	98.18	98.17

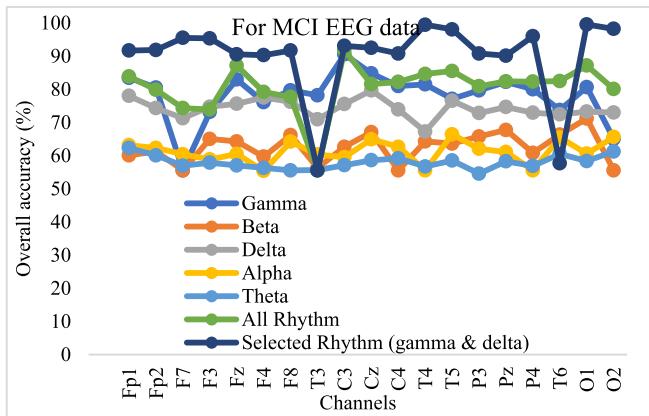


Fig. 16. Performance comparison: Channels vs individual rhythms, all rhythms (together) and selected rhythms (gamma & delta together).

a detailed of MCI dataset. Fig. 16 show overall classification accuracy (in percentage) for each five rhythms, all rhythms (five rhythms together)/selected rhythms (gamma & delta together) against each of the channels. We used the same procedure to choose the best rhythms from the MCI dataset that we did for the AD EEG dataset (explained in the first paragraph of Section B for Table II). As can be seen in Fig. 16, the combination of gamma and delta, called ‘selected rhythm,’ has provided significantly superior results (except in the T3 and T6 channels) than other rhythms. Table IV reports the classification results (in percentage) for the proposed LSTM model for MCI EEG dataset. This table provides the performances of the proposed method for all channels and selected channels with various rhythms. We can see from Table IV that the selected channels with various rhythms give better results compared to all the channels. As seen in Table IV, the selected rhythms (gamma & delta together) in the C3, Fp1, Fz, F8, and Cz channels produce the best results for all the performance measurements. For MCI EEG data, we found that ‘C3, Fp1, Fz, F8, Cz’ channels with gamma and delta band combined are responsible for providing important information for identifying MCI patients from healthy subjects. For AD EEG data, we found that ‘T6, O2, T5, Cz, C3’ channels with the gamma and beta band together conveyed better information for

identifying various categories of AD. In this study, we can see that the biomarkers of MCI detection (responsible rhythms and channels) are a little bit different from AD detection. The reason of this could be that the MCI subjects considered for this study were not turn to AD, as not everyone who has MCI develops AD. The report says that only 10 to 20% of people aged 65 or older with MCI develop AD/dementia over a one-year period [42].

C. Comparison With the Start-of Arts Methods

Table V provides a comparison of our proposed LSTM-based framework against other existing methods that have been developed using the same dataset, that is [14], [38], [39], [40]. This comparison includes the three-class classification problem, that is, NC vs AD1 vs AD2. As seen in Table V, Cassani (2020) et al. [14] introduced a 2-dimensional modulation spectral domain-based feature that achieved overall accuracy of 69.30% for the classification of NC vs AD1 vs AD2. In [38], Cassani et al. employed short-time Fourier transform and continuous wavelet transform (CWT) on resting-state EEG to study the temporal changes of the spectral content of the signals. The obtained features were evaluated by SVM method and acquired an overall accuracy of 64.80% for NC vs AD1 vs AD2 classification. Cassani et al. [39] developed an automated EEG-based AD diagnosis system based on automated artifact removal (AAR) algorithms and a low-density EEG setup. The proposed method achieved an overall accuracy of 91.40%. In [40], Cassani et al. computed same type of features from EEG signals as in [39], and used relevance vector machine (RVM) classifier instead of SVM. The method yielded 84.70% overall accuracy. In [41], Cassani et al. investigated four feature sets (spectral, amplitude modulate rate-of-change, coherence, phase), as well as two combined feature sets (‘All’ and ‘Spectral-modulation’) and tested the effectiveness of the features by SVM classifier. They achieved an average accuracy, 73.50% for the Spectral-modulation feature classifying NC vs AD1 vs AD2. As observed in Table V, all research studies offered for comparison are from the “Cassani et al” group since we included research studies where the same dataset was used as in this study. We discovered some of the work done on this dataset by Cassani et al group in the literature

TABLE V
COMPARISON OF THE PROPOSED LSTM BASED SCHEME WITH THE EXISTING METHODS FOR THE SAME DATASET

Authors	Methods	Classes	Overall Accuracy (%)	EEG datasets
Cassani (2020) et al. [14]	Modulation spectral domain features with SVM classifier	NC vs AD1 vs AD2	69.30	20 cognitively healthy normal control (NC), 19 mild-AD patients (AD1) and 15 patients with moderate-to-severe AD symptoms (AD2)
Cassani (2019) et al. [38]	Short-time Fourier transform, and the continuous wavelet transform (CWT) with SVM	NC vs AD1 vs AD2	64.80	20 cognitively healthy normal control (NC), 19 mild-AD patients (AD1) and 15 patients with moderate-to-severe AD symptoms (AD2)
Cassani (2017) et al. [39]	Spectral features, coherence features, and Amplitude Modulation with SVM	NC vs AD (AD1 & AD2)	91.40	24 healthy elderly, 35 subjects diagnosed with AD
Cassani (2014) et al. [40]	wavelet-enhanced independent component analysis for AAR, three features: spectral power, coherence, and amplitude modulation with RVM classifier	NC vs AD (AD1 & AD2)	84.70	24 cognitively healthy controls, AD comprised 35 mild-to-severe AD patients
Cassani (2014) et al. [41]	Spectral, amplitude modulation rate of change, coherence, and phase with SVM	NC vs AD1 vs AD2	73.50	20 patients with mild AD, 15 with moderate-to-severe AD, and 24 age-matched healthy controls
Our proposed scheme	EEG rhythms and channels-based biomarkers with LSTM architecture	NC vs AD1 vs AD2	97.00	35 NC subjects, 31 AD1 subjects, and 20 AD2 subjects

review. We could not find any other research work for this dataset except Cassani group. Hence, we have just compared our proposed work only with Cassani et al methods.

It is apparent from Table V that our proposed LSTM-based framework achieved the highest accuracy (97%) for AD detection and outperformed all the reported methods that used the same Brazilian EEG AD dataset, thus improving performance from 5.6% to 32.2%. Our proposed LSTM-based framework therefore shows promise as a method for the detection of AD compared to other reported methods.

The current study has some limitations. First, during signal segmentation, the number of windows (such as 16) and window length (such as 100) were chosen based on empirical evidence. Additionally, the hyperparameters for the proposed LSTM method were chosen empirically. Moreover, the two datasets utilized in this study, one comprising 86 participants and the other consisting of 27 participants, are not considered very large in size. We plan to extend the application of this method to clinical huge datasets in the near future.

V. CONCLUSION

This study presents an innovative LSTM-based framework for the efficient identification of different stages of AD (e.g., AD1, AD2) patients from NC subjects by exploring EEG rhythms and channels as biomarkers. A key contribution of this study is the introduction of a new concept to discover responsive biomarkers for AD detection from EEG. The proposed framework was applied to a real-time Brazilian EEG dataset of AD, revealing optimal EEG rhythms and channels that collectively hold significant information for identifying AD categories. The results reveal that the combined gamma and beta rhythms in channels 'Cz, F4, P4, T6, Pz' serve as the best biomarker for AD identification using EEG. The proposed LSTM model shows superior

performance (accuracy 97.00%, sensitivity 96.91%, specificity 98.41%, precision 97.33%, F1 score 97.10%, MCC 95.60%, and Kappa 93.30%) compared to multi-class SVM and the existing methods. Additionally, the proposed LSTM-based model shows excellent results when tested on the MCI EEG dataset, indicating its potential for predicting both AD and MCI. The proposed framework has the advantage of lower experimental costs due to the use of two rhythms and five channels. The framework has the potential to contribute to the development of a CAD system for automatic diagnosis of various AD varieties, leading to improved healthcare and preventing rapid deterioration. In the future, the proposed methodology framework will be extended and implemented for the detection of autism, schizophrenia, and epilepsy from EEG signals.

REFERENCES

- [1] J. Dumurgier and S. Sabia, "Life expectancy in dementia subtypes: Exploring a leading cause of mortality," *Lancet Healthy Longevity*, vol. 2, pp. e449–e450, 2021.
- [2] A. H. Al-Nuaimi et al., "Robust EEG based biomarkers to detect Alzheimer's disease," *Brain Sci.*, vol. 11, 2021, Art. no. 1026.
- [3] B. Klimova, K. Kuca, and P. Maresova, "Global view on Alzheimer's disease and diabetes mellitus: Threats, risks and treatment Alzheimer's disease and diabetes mellitus," *Curr. Alzheimer Res.*, vol. 15, pp. 1277–1282, 2018.
- [4] Causes of death, 2019. [Online]. Available: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/FlagPost/2020/October/Causes_of_death
- [5] National Institute on Ageing, Alzheimer's Disease Fact Sheet. [Online]. Available: <https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet>, Apr. 2023.
- [6] GBD 2016 Mortality Collaborators, "GBD 2016 mortality collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: A systematic analysis for the global burden of disease study 2016," *Lancet*, vol. 390, no. 10100, pp. 1084–1150, Sep. 2017.

- [7] E. Eggink, E. P. Moll van Charante, W. A. van Gool, and E. Richard, "A population perspective on prevention of dementia," *J. Clin. Med.*, vol. 8, no. 6, 2019, Art. no. 834.
- [8] World Health Organisation (WHO), Dementia. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dementia>, Mar. 2023.
- [9] S. Siuly et al., "A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 1966–1976, Sep. 2020.
- [10] B. Oltu, M. F. Akşahin, and S. Kibaroğlu, "A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection," *Biomed. Signal Process. Control*, vol. 63, 2021, Art. no. 102223.
- [11] A. Miltiadous et al., "Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods," *Diagnostics*, vol. 11, 2021, Art. no. 1437.
- [12] S. Nobukawa, T. Yamanishi, S. Kasakawa, H. Nishimura, M. Kikuchi, and T. Takahashi, "Classification methods based on complexity and synchronization of electroencephalography signals in Alzheimer's disease," *Front. Psychiatry*, vol. 11, Apr. 2020, Art. no. 255.
- [13] C. T. Briels, D. N. Schoonhoven, C. J. Stam, H. de Waal, P. Scheltens, and A. A. Gouw, "Reproducibility of EEG functional connectivity in Alzheimer's disease," *Alzheimer's Res. Ther.*, vol. 12, 2020, Art. no. 68.
- [14] R. Cassani and T. H. Falk, "Alzheimer's disease diagnosis and severity level detection based on electroencephalography modulation spectral 'patch' features," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1982–1993, Jul. 2020.
- [15] K. Engedal et al., "The power of EEG to predict conversion from mild cognitive impairment and subjective cognitive decline to dementia," *Dement. Geriatr. Cogn. Disord.*, vol. 49, no. 1, pp. 38–47, 2020.
- [16] X. Bi and H. Wang, "Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning," *Neural Netw.*, vol. 114, pp. 119–135, 2019.
- [17] D. K. Ravikanti and S. Saravanan, "EEG Alzheimer's Net: Development of transformer-based attention long short term memory network for detecting Alzheimer disease using EEG signal," *Biomed. Signal Process. Control*, vol. 86, no. Part C, 2023, Art. no. 105318.
- [18] T. K. K. Ho et al., "DeepADNet: A CNN-LSTM model for the multi-class classification of Alzheimer's disease using multichannel EEG," *Alzheimer's Dement.: J. Alzheimer's Assoc.*, vol. 17, 2022, Art. no. e057573, doi: [10.1002/alz.057573](https://doi.org/10.1002/alz.057573).
- [19] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, and C. Turchetti, "EEG-based Alzheimer's disease recognition using robust-PCA and LSTM recurrent neural network," *Sensors*, vol. 22, no. 10, 2022, Art. no. 3696.
- [20] G. Gkenios, K. Latsiou, K. Diamantaras, I. Chouvarda, and M. Tsolaki, "Diagnosis of Alzheimer's disease and mild cognitive impairment using EEG and recurrent neural networks," in *Proc. IEEE 44th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2022, pp. 3179–3182, doi: [10.1109/EMBC48229.2022.9871302](https://doi.org/10.1109/EMBC48229.2022.9871302).
- [21] A. M. Alvi, S. Siuly, and H. Wang, "A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 2, pp. 375–388, Apr. 2023.
- [22] S. Sridhar and V. Manian, "EEG and deep learning based brain cognitive function classification," *Computers*, vol. 9, no. 4, 2020, Art. no. 104.
- [23] A. A. Petrosian, D. V. Prokhorov, W. Lajara-Nanson, and R. B. Schiffer, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clin. Neurophysiol.*, vol. 112, pp. 1378–1387, 2001.
- [24] S. M. D. Brucki, R. Nitrini, P. Caramelli, P. H. F. Bertolucci, and I. H. Okamoto, "Suggestions for utilization of the mini-mental state examination in Brazil," *Arquivos Neuro-Psiquiatria*, vol. 61, no. 3B, pp. 777–781, 2003.
- [25] G. D. D. McKhann, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.
- [26] F. J. Fraga et al., "Towards an EEG-based biomarker for Alzheimer's disease: Improving amplitude modulation analysis features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 1207–1211.
- [27] R. Cassani, T. H. Falk, F. J. Fraga, P. A. M. Kanda, and R. Anghinah, "The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis," *Front. Aging Neurosci.*, vol. 6, 2014, Art. no. 55.
- [28] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," *J. Neurosci. Methods*, vol. 158, no. 2, pp. 300–312, Dec. 2006.
- [29] P. Ghorbanian, D. M. Devilbiss, A. J. Simon, A. Bernstein, T. Hess, and H. Ashrafiun, "Discrete wavelet transform EEG features of Alzheimer's disease in activated states," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2012, pp. 2937–2940.
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [31] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3113–3122, Mar. 2020.
- [32] U. Budak, V. Bajaj, Y. Akbulut, O. Atilla, and A. Sengur, "An effective hybrid model for EEG-based drowsiness detection," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7624–7631, Sep. 2019.
- [33] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, pp. 115–143, Aug. 2002.
- [34] D. Arifoglu and A. Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks," *Procedia Comput. Sci.*, vol. 110, pp. 86–93, 2017.
- [35] S. Siuly and Y. Li, "A novel statistical algorithm for multiclass EEG signal classification," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 154–167, 2014.
- [36] M. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG," *PLoS One*, vol. 16, no. 6, 2021, Art. no. e0253094.
- [37] S. Siuly, S. 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.
- [38] R. Cassani and T. H. Falk, "Automated Alzheimer's disease diagnosis using a low-density EEG layout and new features based on the power of modulation spectral 'patches,'" in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2019, pp. 1259–1263.
- [39] R. Cassani, T. H. Falk, F. J. Fraga, M. Cecchi, D. K. Moore, and R. Anghinah, "Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices," *Biomed. Signal Process. Control*, vol. 33, pp. 261–271, 2017.
- [40] R. Cassani, T. H. Falk, F. J. Fraga, P. A. Kanda, and R. Anghinah, "Towards automated EEG-based Alzheimer's disease diagnosis using relevance vector machines," in *Proc. IEEE 5th ISSNIP Biosignals Biorobotics Conf.: Biosignals Robot. Better Safer Living*, 2014, pp. 1–6.
- [41] R. Cassani, T. H. Falk, F. J. Fraga, P. A. M. Kanda, and R. Anghinah, "The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis," *Front. Aging Neurosci.*, vol. 6, 2014, Art. no. 55.
- [42] National Institute on Ageing. [Online]. Available: <https://www.nia.nih.gov/health/what-mild-cognitive-impairment>, Apr. 2021.
- [43] M. M. Mukaka, "Statistics corner: A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, Sep. 2012.



Siuly Siuly (Member, IEEE) received the Ph.D. degree in biomedical engineering from the University of Southern Queensland, Toowoomba, QLD, Australia, in September 2012. She is currently a Research Fellow with the Institute for Sustainable Industries and Liveable Cities, Victoria University, Footscray, VIC, Australia. She already made significant contributions to the mentioned areas that have been published in top leading journals and conferences. Her research interests include biomedical signal processing, analysis and classification, detection and prediction of neurological abnormality from EEG signal data, artificial intelligence, medical data mining and brain-computer interface. Dr. Siuly is an Associate Editor (AE) for IEEE TRANSITION ON NEURAL SYSTEMS & REHABILITATION ENGINEERING, Managing Editor of *Health Informatics Sciences and Systems* and Academic Editor of *PLOS One* and an AE for *Frontiers in Medical Technology*.



Ömer Faruk Alçin received the bachelor's degree in electronics and computers education, the master's degree in electronics education, and the Ph.D. degree in electrical-electronics engineering from Fırat University, Elazığ, Turkey, in 2007, 2011, and 2015, respectively. In February 2009, he became a Research Assistant with the Technical Education Faculty, Fırat University. He is currently an Assistant Professor with the Electrical Engineering Department, Malatya Turgut Özal University, Malatya, Turkey. His research interests include signal processing, machine learning, medical image processing, and deep learning.



Yan Li (Senior Member, IEEE) received the Ph.D. degree from the Flinders University of South Australia, Bedford Park, SA, Australia. She is currently a full Professor of computer science and the Associate Head (Research) with the School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD, Australia. She has authored or coauthored more than 190 high quality publications and recently obtained more than \$2.2 million research grants from Australia government and through international collaborations. She is the Leader of USQ Data Science Programs. Her research interests include artificial intelligence, machine learning, Big Data and Internet technologies, signal/image processing and EEG research. Dr. Li was the recipient of many research and teaching excellence awards, including 2012 Australia prestigious National Learning and Teaching Citation Award, 2008 Queensland Government Smart State-Smart Women Award, 2009 USQ Teaching Excellence Award, 2009 USQ Research Excellence Award, and 2015-2018 Research Publication Excellence Awards.



Hua Wang (Senior Member, IEEE) received the Ph.D. degree from the University of Southern Queensland, Toowoomba, QLD, Australia. He was a Professor with the University of Southern Queensland. He is currently a full-time Professor with Victoria University, Melbourne, VIC, Australia. He has more than ten years teaching and working experience in applied informatics with both enterprise and university. He has expertise in electronic commerce, business process modeling, and enterprise architecture. As an Chief Investigator, three Australian Research Council Discovery grants have been awarded since 2006, and 280 peer reviewed scholar papers have been published. Ten Ph.D. students have already graduated under his principal supervision.



Peng (Paul) Wen (Senior Member, IEEE) received the Ph.D. degree from the Flinders University of South Australia, Bedford Park, SA, Australia. He is currently a full Professor and acting Head of the School of Engineering, University of Southern Queensland, Toowoomba, QLD, Australia. He has more than 170 research articles in leading international journals and conferences and grant more than 3.4 million. His research interests include biomedical engineering, complex medical engineering, networked system, intelligent control. Prof. Wen is a Member of IEEE, IEEE Engineering of Medicine and Biology, and Institute of Complex Medical Engineering.