Continuous Scoring of Depression from EEG Signals via a Hybrid of Convolutional Neural Networks

S. Hashempour, R. Boostani, M. Mohammadi, and S. Sanei, *Senior Member, IEEE*

***Abstract*— Depression score is traditionally determined by taking the Beck depression inventory (BDI) test, which is a qualitative questionnaire. Quantitative scoring of depression has also been achieved by analyzing and classifying pre-recorded electroencephalography (EEG) signals. Here, we go one step further and apply raw EEG signals to a proposed hybrid convolutional and temporal-convolutional neural network (CNN-TCN) to continuously estimate the BDI score. In this research, the EEG signals of 119 individuals are captured by 64 scalp electrodes through successive eyes-closed and eyes-open intervals. Moreover, all the subjects take the BDI test and their scores are determined. The proposed CNN-TCN provides mean squared error (MSE) of 5.64±1.6 and mean absolute error (MAE) of 1.73±0.27 for eyes-open state and also provides MSE of 9.53±2.94 and MAE of 2.32±0.35 for the eyes-closed state, which significantly surpasses state-of-the-art deep network methods. In another approach, conventional EEG features are elicited from the EEG signals in successive frames and apply them to the proposed CNN-TCN in conjunction with known statistical regression methods. Our method provides MSE of 10.81±5.14 and MAE of 2.41±0.59 that statistically outperform the statistical regression methods. Moreover, the results with raw EEG are significantly better than those with EEG features.**

***Index Terms*— Beck depression test, CNN, EEG, TCN, Deep Learning.**

I. Introduction

DEPRESSION, with more than 264 million individuals involved globally, is a significant public health issue that extensively exerts influence on people's quality of life [1]. This mental disorder encompasses various physical and mental manifestations, including sleep disruption, low self-esteem, discouragement, appetite changes, poor concentration, and in chronic cases, suicide ideations [2]. Early diagnosis of depression is crucial for more effective treatment [3], [4]. The revised Beck Depression Inventory (BDI-II) is frequently used for depression screening by specialists [5]. This test contains 21 questions, which aim to evaluate the feedback and symptoms of depressed patients. The score of this test can be discretely varied from 0 to 63 [6]. It should be pointed out that the BDI-II test is qualitative and does not stem from strong physiological basis. Since depression affects the secretion of neurotransmitters in the human brain, it is logical to expect that it influences the neurons' electrical activity, recorded using electroencephalography (EEG). This signal, records the detailed physiological functions of the brain and contains rich temporal information that can be decoded and interpreted with various conventional feature extraction techniques. EEG analysis is used for the diagnosis of different brain diseases like bipolar manic depression (BMD) [7], [8], seizure [9], sleep disorder [10], schizophrenia [11], anxiety [12], Alzheimer's [13].

To differentiate between the patients with depression and normal subjects, Subha et al. [14] estimated several EEG features like relative wavelet energy (RWE), sample entropy and applied them to a two-layer feedforward artificial neural network (ANN). They reported 98.11% classification accuracy. Ahmadlou et al. [15] presented a wavelet-chaos methodology for characterizing the EEG signals captured by the scalp electrodes located on the frontal brain lobe of patients diagnosed with major deep depression (MDD). They applied Higuchi and Katz fractal dimension to an enhanced probabilistic neural network classifier. They achieved 91.3% accuracy to distinguish MDDs from the healthy subjects. In another attempt, Ahmadlou et al. [16] extracted nonlinear EEG features, selected the most discriminative ones and applied this subset of features to logistic regression to categorize healthy subjects from depressed ones. Faust et al. [17] extracted both wavelet packet coefficients and entropy features from the EEG of depressed patients and control subjects. The selected features were classified using the probabilistic neural network classifier, which yielded 98.20% and 99.50% accuracy for the scalp electrodes located on the left and right hemispheres, respectively. Acharya et al. [18] extracted several nonlinear features such as detrended fluctuation analysis (DFA), fractal dimension, higher-order spectra (HOS), Hurst's exponent (HE), largest Lyapunov exponent (LLE), recurrence quantification analysis (RQA), and sample entropy (SampEn) from the EEGs of controls and depressed subjects. The estimated features were ranked according to their significance and fed into five different classifiers. They reported an average accuracy of 98% using a support vector machines (SVM) classifier. Liao et al. [19] utilized eigen-filter-bank common spatial patterns (CSP) to extract Spatio-temporal EEG features and then applied

S. Hashempour and R. Boostani are with CSE&IT Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran. (e-mail. hashempour.sara73@gmail.com and Boostani@shirazu.ac.ir).

M. Mohammadi is with the Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, KR-Iraq. (e-mail: mukhtar@lfu.edu.krd).

S. Sanei is with School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK (e-mail. saeid.sanei@ntu.ac.uk).

principal component analysis (PCA) to achieve discriminative features. The features were applied to an SVM, which led to the depression diagnosis accuracy of 81.23%. Bairy et al. [20] used an EEG-based computer-aided diagnosis (CAD) system for the diagnosis of depression based on a linear predictive coding (LPC) scheme by using HOS parameters to extract significant features for classifying depressed from non-depressed subjects. They obtained sensitivity, specificity, and accuracy of 91.46%, 97.45%, and 94.30%, respectively.

In a different approach, Acharya et al. [21] tried to differentiate depressed patients from controls by applying their raw EEG signals to a 13-layer convolutional neural network (CNN). They attained 93.5% and 96% classification accuracy using the scalp electrodes located over the left and right hemispheres. Li et al. [22] focused on designing a CAD system using CNN to detect mild depression. They reported an accuracy of 85.62% for the mild depression diagnosis. Saeedi et al. [23] developed a deep learning framework comprised of sixteen connectivity methods in eight EEG frequency bands to detect MDD patients automatically. They introduced the extracted images of connectivity from EEG signals and applied them to a convolutional neural network, long short-term memory (LSTM) network, and a combined CNN-LSTM model. They achieved the highest accuracy of 99.24, the sensitivity of 100, and the specificity of 99.25 using the CNN-LSTM network.

To the best of the authors' knowledge, no study has been conducted to estimate the Beck score of subjects by analyzing their EEG signals. In this regard, we have used a rich dataset containing the EEG signals of 119 patients, all of whom have executed the Beck test. To estimate the depression score, we have deployed two approaches. In one approach, raw EEG signals are applied to the proposed combinatorial convolutional and temporal convolutional neural network (CNN-TCN) to assess the Beck score. In another approach, we have elicited informative features from their scalp EEGs and feed them to the proposed CNN-TCN for estimating their Beck score.

The remainder of this paper is organized as follows. Section II introduces the characteristics of the deployed dataset and then proposes the suggested candidate features and the deep hybrid network. Section III illustrates the empirical results and their merits and demerits with the results of the compared methods. Finally, Section IV concludes the paper.

II. MATERIALS AND METHODS

A. Dataset and Preprocessing

We have used a publicly available dataset on PRED+CT website [24], originally containing EEG signals of 121 subjects (72 females and 49 males, 18.86±1.19), among whom two subjects' practical information were missed and removed accordingly [25]. The subjects have different depressive levels. Among these enrolled participants, the Beck score of 76 subjects was in the range of 0-13, categorized as the control group (without depression). The score of 14 subjects was in the interval of 14-19 (mild depression), the score of 24 subjects was in the range of 20-28 (moderate depression), and the score of 5 subjects ranged from 29-63, which is considered as severe depression [26]. The data comprised 500 seconds of recorded signals via 64 channels with electrode settings according to 10-20 standard EEG recording system and sampling frequency of 500 Hz during the resting state. The paradigm recording for subjects included eyes-opened and eyes-closed events with varied sizes. All participants provided written consent approved by the University of Arizona approves. Subjects' age is in the range of 18–25 years old, and are not having any history of head trauma or seizures. They are not taking any psychoactive medications. Participants are enrolled from introductory psychology categories based on their BDI scores in a mass survey.

To preprocess the EEG signals in this work, first, the baselines of the signals are removed. Then they are passed through a notch filter of 50 Hz [27], [28] for eliminating the power grid effect. Subsequently, the signals are passed through a bandpass filter with the cut-off frequencies of 0.2 and 50HZ and following this, a Butterworth filter (order=5, high cut=50HZ, low cut= 1HZ) was carried out [2]. In the last step, the independent component analysis (ICA) is applied to the filtered EEGs for eliminating the remaining undesired components. This research utilizes the MNE-python package, which uses a semi-automated ICA approach for parsing the contaminations. Here, fastICA has been employed, which is considerabely faster than conventional ICA approaches and maximizes the non-Gaussianity. In this method, the mixtures are whitened by PCA and then decomposed by ICA. Afterwards, the artifacts are detected via MNE. Then the remaining ICA components are back-projected to the channel space [29], [30].

To process the raw data, EEG data is transformed into two separate datasets: eyes-open and eyes-closed resting-state datasets. The difference between the number of samples for each subject and the memory size limitations enforced the following steps. Firstly, we down-sampled the EEGs by a factor of two to avoid approaching the Nyquist rate as well as reducing the volume of input data**.** Secondly, two minutes (3000 samples) of each subject's EEG signals in eyes-open and eyes-closed states are selected. The signals are segmented in the interval of five seconds (1250 samples), where successive windows have 90% overlap. Afterwards, for achieving more consistent estimations, we balance the data, based on the number of subjects, over the entire range of Beck scores. This imbalanced problem stems from the fact that the sampling probabilities vary across depression levels, which means that the data is not uniformly distributed. Therefore, in each window, the number of data points in depression classes is equalized to the class with the lowest size (number of data points).

B. Feature Extraction

In the feature-based approach, for each subject, the EEG data is segmented event-wise. Afterwards, the following features are estimated for each individual's signal [31]. We compute the features in four domains: (I) time-domain features [28], [29], including min, max, standard deviation, mean, median, activity, mobility, complexity, kurtosis, 2^{nd} difference

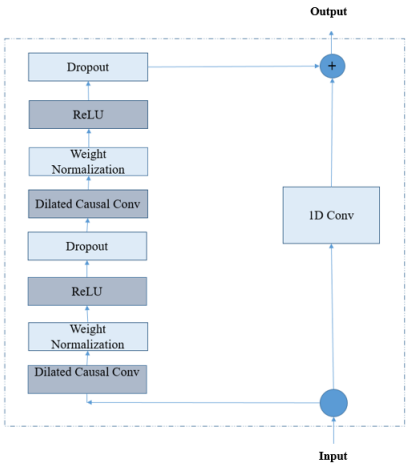


Fig. 1: Temporal convolutional network (TCN) residual block.

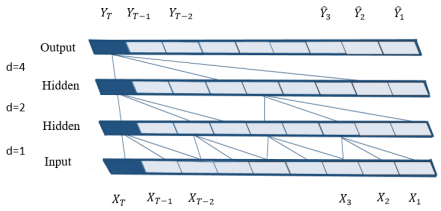


Fig. 2: Dilated causal convolution with dilation factors, d= 1, 2, 4, and filter size =3.

mean, 1^{st} difference max, coefficient of variation, skewness; (II) wavelet domain features [32], [33] including mean, standard deviation, and energy of wavelet approximation and wavelet details; (III) complexity-based features [34] (mean of the vertex to vertex slope, the variance of the vertex to vertex slope); (IV) frequency domain features (max power spectrum in the Delta, Theta, Alpha, and Beta bands, and the ratio of Delta/Alpha [34], [35] and Delta/Theta band powers).

C. The proposed hybrid deep CNN-TCN network

The EEG signal can be viewed as a temporal series of brain activity signals. EEG data collected over a period of time are used to identify a variety of disorders. The suggested deep network for both raw EEG data and elicited EEG features is a cascade combination of CNN being capable of extracting local characteristics and TCN which can efficiently be employed for sequence modeling tasks by storing exceptionally lengthy historical data to estimate the depression (Beck) score. CNN's are potent tools for feature learning, classification, and regression. In general, CNNs are comprised of three main layers: (I) convolutional layer, which performs the feature extraction procedure via different linear and nonlinear functions (II) pooling layer for reducing the spatial size of the feature map by down-sampling process (III) fully connected layer in which the regression task is performed by flattening the outputs of previous layers into a single vector. In addition, some more parameters such as the number of filters, kernel size, stride, and padding need to be appropriately tuned for customizing a CNN. CNN, despite its excellent feature extraction strength, struggles to preserve time coherence. Although networks such as RNN and LSTM have been proposed to address the matter of high-dimensional time-series data, they have drawbacks such as the disappearance of gradient and higher running time.

TCN is an extension of CNN, which is used in the temporal sequence modeling tasks and has shown better performance than canonical recurrent networks such as long short-time memory (LSTMs) across a broad range of datasets and tasks. [36], [37]. TCN utilizes causal 1D convolutions, dilated convolutions, and residual layers. Combining dilated convolutions and residual blocks results in high receptive field size and addresses the computational complexity problem. Moreover, using Causal convolutions in the architecture is to prevent data from being transported from the future to the past and to ensure that the output sequence is the same length as the input sequence; Zero-padding is used [38]. Unlike the standard convolution layer, which can only look back at history with size being linear with the depth of the network, TCN using dilated convolution layers results in an exponential enlargement of the receptive field size. In this layer, a filter is applied over a region more significant than its size by skipping input data with a specified step similar to pooling or stride convolutions since it increases the size of the receptive field, but the output is equal to the input. 1D dilated convolution operation for an input sequence $x \in R^T$, and a filter $h = \{0, ..., k-1\}$ is defined as:

$$H(s) = (x *_d h)(s) = \sum_{i=0}^{k-1} h(i).x_{s-d\times i} \quad (1)$$

where $*_d$ represents the convolution operation with dilation factor $d = 2^v$, v is the network level, k is the filter size, and the term $s - d \times i$ demonstrates the direction of the past. Increasing the dilated factor d exponentially with the network depth when using dilated convolutions, ensures the full history coverage by the receptive field. Figs. 1 and 2 show the architectural components of the TCN model. As demonstrated in Fig. 1, in the main path of the residual block [39] there are two dilated casual convolution layers followed by a nonlinear rectified linear unit (ReLU) layer [40], spatial dropout [41], and weight regularization layers for generalization. Augmenting a residual block to a TCN capacitates the receptive field twice Because it has two basic causal convolutional layers. The residual block' output, computed by adding the input x to the series of transformations F on this input, is presented in the following equation.

$$O = activation(x + F(x)) \quad (2)$$

Dilated convolution demonstrated in Fig. 2. is used to solve the small receptive field size problem. The TCN block comprises several dilated convolutions with an input which is a coefficient equal to the powers of two (incrementally).

In the obtained dataset, each subject has about 500 seconds of recorded EEG signals captured by 62 proper scalp channels in addition to two more channels HEOG and VEOG. The last two channels were removed in the preprocessing phase.

In feature-based methodology, the input of the proposed CNN-based network would be in the form of a matrix. In the first step, each individual's EEG signal is segmented

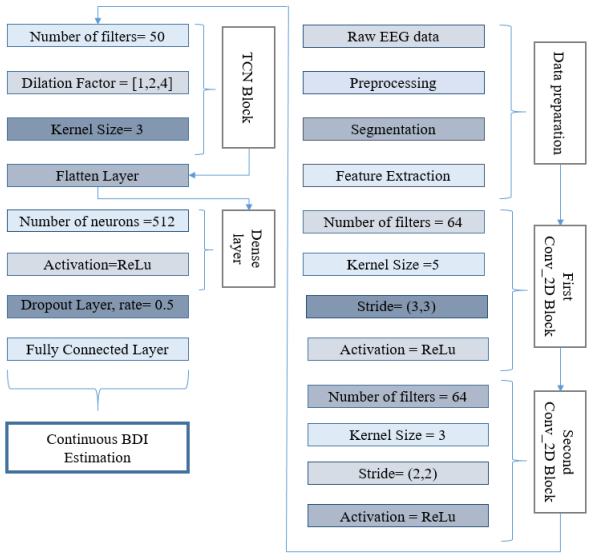


Fig. 3: The proposed network structure in the high-level presentation.

at event points.The signal comprises 12 unique events, and since each subject's EEG data contains a variable number of events from each distinct event, the number of segments formed for each specific event differs. To achieve a more balanced data, we equalize the number of sections regarding each unique event to the minimum number of segments, which is 120. As a result, the data structure for each subject becomes $(12, 120, k, 62)$, where 12 represents the number of unique events, 120 represents the number of segments, k represents the length of the signal data in each segment, and 62 is the number of channels in each recorded signal. Subsequently, 31 linear and nonlinear features are extracted in each channel of the segmented EEG signals. In this step, an array of (12, 120, 62, 31) is created for each individual, and since the last two dimensions' types are features, they are concatenated. The network's input matrix is generated by stacking the created data for all 119 subjects in the dataset, leading to the final multi-modal structure of $((119 \times 12), 120, (62 \times 31))$.

To feed the features into the first convolutional network (Conv2D, stride= (1,7)), we have to add another dimension at the end of the input data with the value of 1 because the input for the Conv2D should be a four-dimensional matrix. We concatenate the second and third dimensions for dimension reduction as they are both features; consequently, the size of input data would be $((119 \times 12), 120, 1992, 1)$. This layer is connected to another Conv2D (stride= (1,3)) layer connected to the TCN block, inspired by Bai et al. [42]. Subsequently, we use a dropout layer to avoid over-fitting and improve the generalization of the model. This layer is connected to the final fully-connected layer in which the regression is performed. We use a ReLU function after each convolution layer and the first dense layer. This function takes the value 0 for the negative inputs and the x value for the positive inputs x. Table I presents the detailed information and parameter settings of the proposed methodology, and conventional machine learning methods parameters are shown in Table II. Parameters are set by trying various network structures to avoid over-fitting and reduce regression errors as much as possible. The high-level information and parameters of the proposed design are mentioned in Fig. 3. For weight initialization of the proposed network, Glorot uniform initialization is utilized, which is a common initialization scheme for deep neural networks. The overall algorithm of the proposed network id presented in Algorithm 1. The same network has been obtained for the raw EEG data analysis (eyes-open and eyes-closed data separately) aiming for the BDI prediction. The source code is available at: https://github.com/HashempourSara/depression-score-estimation

Algorithm 1 Algorithm of the proposed model

Input: D: EEG signal of patients, L: BDIs of patients
Output: Trained Neural Network
Step 1: Data Preparation
X = [ ]; Y = [ ]
for each patient signal in D and patient BDI in L **do**
  Split eyes open and eyes close sequence data
  for each open/close sequence data **do**
    Remove baseline
    Notch filter
    Band-pass filter
    Butterworth filter
    ICA
    EEG sequence windowing
    Add created windowed signals to X
    Add patient BDI with the number of created windows to Y
  end for
end for
Step 2: BDI prediction
while not convergence **do**
  Divide X, Y into b mini batches of size $i = 1$ to b
  Choose the i^{th} mini batch of X
  Compute the output of network
  Compute the errors (*with Table III*)
  Update the network weights
end for

TABLE I: THE PROPOSED MODEL PARAMETERS AND SETTINGS.

Parameters Setting	values
Batch size	64
Optimizer	Adam
Metric	MSE, RMSE, MAE,R^2
Learning Rate	0.0001
Epochs	50

TABLE II: THE SETTINGS FOR CLASSICAL MACHINE LEARNING METHODS.

Classical methods	Parameters
KNN	Number of neighbors=5
SVR	C=100, kernel= rbf, max iter=-1
RF	N_estimators= 100, max_depth=3, n_jobs=-1

D. Compared Methods

When the proposed network works with raw EEGs, we compare it with the three successful deep learning-based

TABLE III: THE REGRESSION METRICS FORMULAS

Formula	Regression Metrics
$1/n\sum_{i=1}^{i=n}(y_i-\hat{y}_i)^2$	MSE (Mean squared error)
$\sqrt{1/n\sum_{i=1}^{i=n}(y_i-\hat{y}_i)^2}$	RMSE (Root mean square error)
$1/n\sum_{i=1}^{i=n}\lvert y_i-\hat{y}_i\rvert$	MAE (Mean absolute error)
$\sum_{i=1}^{i=n}({y_i}^2/\hat{y}_i^2)$	Rsquared

models, including the combination of convolutional neural network and long short-term memory (CNN-LSTM) proposed by Ay et al. [43], a 13 layer deep CNN offered by Achariya et al. [21] and CNN based model introduced by Li et al. [22].

Another comprehensive comparison of the proposed CNN-TCN method with the popular statistical machine learning-based methods like support vector regression (SVR) [44], K-nearest neighbor [45], and random forest [46] is performed by applying the EEG features to them for predicting the depression severity score. The feature selection methodology used for the classical machine learning methods is minimum redundancy maximum relevance, in which the features are selected based on being the most relevant regarding the target and the least redundant for the previously chosen features.

In the K-fold cross-validation process, a model is trained using $(k-1)$ folds as training data and tests it over the left fold. This procedure is repeated k times until all folds have a chance to be selected as the test set. The final result is the average of the k evaluations. A 10-fold cross-validation procedure is utilized for model evaluation using 10% of the subjects as a testing set [47]. The overall performance is computed by averaging the results from all ten assessments.

Standard statistical criteria for evaluate the methods used in estimating the severity of depression by regression methods have been used, including mean squared error, root mean square error, mean absolute error, and R squared.The description and the formula of these measurements are presented in Table III, in which y and ŷ represent the actual and predicted values.

III. Results and Discussion

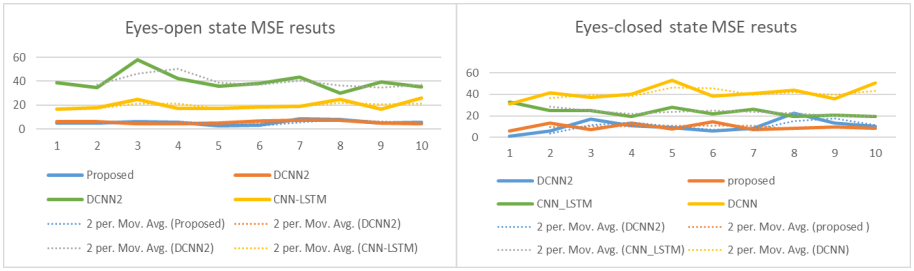
In this section, the depression score is determined in two different manners: using raw EEGs and estimated EEG features. As we mentioned before, just deep learning methods are compared in the former manner, while in the latter mode, the proposed deep learning is compared to SVR, KNN, and random forest methodologies. Finally, the overall performance in both states is computed by averaging the results over ten folds. Normalization operations and window-wise balancing for the raw data are performed to circumvent imbalanced data distribution.

In the presented CNN-TCN structure, for the raw data manner, discriminant features are extracted by the convolutional layers and then feed these feature vectors to the TCN block to perform the regression task. In the feature-based approach, the secondary feature extraction is performed by the CNN blocks and, following this, would be transformed into the TCN layers for BDI estimation. As a sequence modeling manner, one of the main limitations of TCN is that it may not be suitable for processing temporally large datasets. Consequently, in the proposed network, using 2D convolutions in the first two layers, the network time steps and the number of features are reduced before feeding them into the TCN block. This reduction of features also accelerates the processing procedure. As a result, the network might be less time-consuming and contains fewer data samples leading to less complexity.

The moving average trends of MSE results per 10 folds adopting deep learning-based networks can be compared in Fig. 4, and as depicted in these graphs, the CNN-TCN results surpass other methods. The overall results gained by the feature-based CNN-TCN methodology and state-of-the-art approaches are reported in terms of MSE, RMSE, *R* squared, and MAE in Table IV. Furthermore, the results are enhanced using the frame-wise raw-EEG-based regression prediction, which is presented in Table V. Although the achieved results by the proposed model and CNN-LSTM model are fairly close to each other, the relative superiority of the introduced model is that TCN structures perform better in modeling long-term dependencies. It has a more extended adequate memory size in comparison with LSTMs. Furthermore, to examine the effect of the number of channels, the results of the deep learning models are also compared using 14 and 32 EEG channels, which is presented in Table VI. Additionally, regression results for sampling frequencies of 250 and 1000 are depicted in Fig. 5 to survey the impact of the varied sample rates.It can be understood from this figure that number of samples affects the network's learning process, and the more the sampling frequency is, the better results are achieved. The intra-subject coefficient of variation (CoV) is computed in this study, which is an average of the CoV in the error of predicted BDI values calculated from windows of each individual's EEG data. The inter-subject CoV is an average value calculated from the 10-times-10-folds error of predictions for all subjects. The intra-individual and the inter-individual coefficient of error variability for predicted BDI scores is 0.45, and 0.82 respectively, indicating room for further improvement of the results. It should also be noted that the proposed method is offline in the training phase, and after the training procedure the estimation of BDI values can be performed online.

A. Computational Complexity

The computational complexity of the proposed model with the other three deep learning-based methods is also compared, presented in Table VII, in terms of the number of flops (floating-point operations per second), inference speed (number of examples per second), and the average training time in 10 folds [48]. The CNN-TCN network superseded the models by Li et al., 2019 and Ay et al., 2019 regarding the complexity measurements by having less run time, fewer flops, and more speed. This is because the LSTM network's training pace is significantly slowed by non-parallelizability, which consumes excessive computational resources. TCN, on the other hand, solves this difficulty by utilizing a backpropagation



(a) (b)

Fig. 4: The comparison between the MSE results and moving average trends of deep models per fold using the raw a) eyes-open state EEG data, and b) eyes-closed state EEG data.

path independent of the sequence's temporal direction and provides a unified approach that hierarchically captures low and high-level temporal information with fewer complexities and a larger receptive field size. Although the deep network presented by Acharya et al., 2018 were less complex than our represented model, the superiority of the results obtained by CNN-TCN is considerable compared to those of the model proposed by Achariya et al. Moreover, the results achieved by Li et al., 2019 network were moderately close to the presented model results, while having higher computational complexity.

B. Robustness

Diverse network configuration modes using TCN structure are regulated. Ultimately, the best model parameters achieving the highest performance are used for this study through trial-and-error by optimizing the model's error. Training and test curves are examined to avoid over-fitting, and the most appropriate parameters are selected. Furthermore, white noises with different intensities (SNRs of 10, 20, 30, and 40 dB) are added to the signal to show the reliability of the network structure, and then the network results for the eyes-open state are reported accordingly. As represented in Table VIII, the differences in the regression results can be considered minor, manifesting the robustness of the model versus the increasing noise level.

C. Contribution

The model structure introduced in this research is the main contribution of this article. This is because the number of layers and the combination of CNN and TCN networks is a new structure. The structure of TCN solely was also tested for this study, which yielded poorer results than the combinatorial model. We have utilized Conv2D layers in the primitive layers of the model due to their outstanding ability in extracting discriminant features. After rigorous examinations, we proved that a combined CNN-TCN approach outperforms TCN.

In this study, for the first time, the BDI index, being a valid criterion for detecting the severity of depression, is estimated with a very small error. To the best of our knowledge, most studies in the field of EEG-based depression diagnosis mainly focus on the classification of normal vs. depressed subjects or MDD diagnoses. We have used a rich dataset containing individuals with mild, moderate, and severe depression. The best network parameters have been derived by try and error.

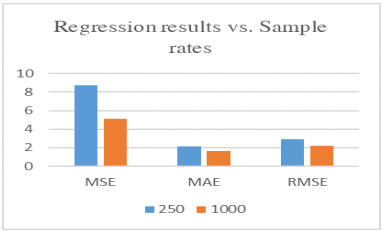


Fig. 5: Model results with twice and half times the default data sampling rate.

IV. Conclusion

In this research, we explored both conventional machine-learning models and the temporal CNN-based deep model to predict the severity of depression in terms of BDI continuously. With the change in depression recognition area, both spectral and temporal information substantially vary. As a result, we use the CNN structure with its outstanding feature extraction ability and combine it with the TCN block, a powerful learner with a high spectral resolution. The presented methodology is developed utilizing both raw and features extracted from the EEG data of 119 individuals. The experimental results are surveyed via subject-wise 10-fold cross-validation demonstrate that the CNN-TCN methodology outperforms other benchmarks in BDI detection of depressive subjects, which can aid the specialists in early depression detection and better treatments.The suggested robust model can be utilized for EEG-based detection of other neurological diseases too. However, one of the significant challenges in this work is due to the inequality of the data size in different depressive states leading to an imbalanced data problem. The proposed model has its limitations. To begin with, TCN may require ample data storage during evaluation compared to CNN networks since it needs to involve a reasonable size data history. Also, for the hyper parameter tuning, methods such as grid search, random search, and Bayesian optimization are more convenient than the many time-consuming trial-and-error approaches we have used. The deep learning models are black boxes. Therefore, interpreting the the network's function is important. For a more explanatory model, a regression activation map can be generated to explain the regression results. As part of our future work, we aim to progress the proposed methodology in the following ways to achieve

TABLE IV: PRESENTATION OF THE REGRESSION RESULTS USING THE FEATURE-BASED APPROACHES.

Approaches	MSE± SD	RMSE± SD	MAE± SD	R Squared± SD
KNN	121.96±10.4	11.03±0.47	7.51±0.44	23.12+0.074
SVR	87±3.77	9.32±0.2	7.4±0.17	19.18±0.03
Random Forest	79.52±2.66	8.51±0.15	7.31±0.16	32.25±0.023
Achariya et al., 2018	58.11±18.1	7.54±1.08	5.27±0.46	46.1±0.16
Li et al., 2019	29.81±14.89	5.3±1.28	3.81±1.01	72.25±0.14
Ay et al., 2019	25.09±5.1	4.98±0.49	10.88±5.14	76.68±0.047
Proposed feature-based model	10.81±5.14	3.21±0.73	2.41±0.59	90.26±0.046

TABLE V: COMPARISON OF THE REGRESSION RESULTS USING THE RAW DATA.

Approaches	Data acquisition state	MSE± SD	RMSE± SD	MAE± SD	R Squared± SD
Proposed model	Eyes-open	5.64±1.6	2.34±0.36	1.73±0.27	94.95±0.015
	Eyes-closed	9.53± 2.94	3.05± 0.46	2.32± 0.35	91.47± 0.02
Achariya et al., 2018	Eyes-open	39.56±7.15	6.26±0.54	4.97±0.5	64.62± 0.064
	Eyes-closed	41.22± 6.26	6.4± 0.48	4.92± 0.36	63.14±0.056
Li et al., 2019	Eyes-open	5.96 ± 1.14	2.43 ± 0.23	1.8 ± 0.16	94.67 ± 0.01
	Eyes-closed	10.88± 5.14	3.21± 0.73	2.41± 0.59	90.26 ± 0.046
Ay et al., 2019	Eyes-open	20.02± 3.53	4.45± 0.38	3.44± 0.34	82.09± 0.031
	Eyes-closed	23.69±4.19	4.84±0.42	3.76±0.39	78.82± 0.037

TABLE VI: COMPARISON OF THE RESULTS USING DIFFERENT NUMBER OF EEG CHANNELS.

Network, channels	MSE	RMSE	MAE	R Squared
Proposed model- 32 channels	11.43±4.52	3.31±0.65	2.4±0.048	89.77±0.04
Proposed- 14 channels	23.67±8.77	4.78±0.89	3.57±75.52	78.84±0.078
Li et al., 2019- 32 channels	14.6±4.74	3.76±0.64	2.89±0.45	86.93±0.04
Li et al., 2019- 14 channels	26.22±6.24	5.09±0.59	3.92±0.43	76.49±0.05
Achariya et al., 2018- 32 channels	49.51±8.78	7.00±0.62	5.47±0.44	55.74±0.078
Achariya et al., 2018- 14 channels	53.83±5.50	7.32±0.37	5.91±0.34	51.87±0.049
Ay et al., 2019- 32 channels	26.31±4.16	5.11±0.42	3.93±0.39	76.48±0.036
Ay et al., 2019- 14 channels	53.83±5.50	7.32±0.37	5.91±0.34	51.87±0.049

TABLE VII: COMPUTATIONAL COMPLEXITY OF THE DEEP LEARNING-BASED NETWORKS.

Model	Flops	Training time	Inference speed
Achariya et al., 2018	129746	313.84	1654.34
Ay et al., 2019	41208008	2915.29	129.42
Li et al., 2019	3054482	4322.43	91.46
proposed model	10838963	339.93	874.71

TABLE VIII: COMPARISON OF THE PROPOSED MODEL PERFORMANCE FOR BDI PREDICTION BY ADDING GAUSSIAN NOISES AT DIFFERENT LEVELS TO PROVE THE RELIABILITY OF THE NETWORK.

SNR	MSE± SD	MAE± SD	RMSE± SD	R Squared± SD
10	5.84±2.82	1.71±0.43	2.27±0.37	94.1±0.025
20	6.28±2.2	1.78±0.3	2.47±0.41	94.38±0.019
30	6.71±2.44	1.86±0.34	2.54±0.45	94±0.021
40	8.82±3.54	2.13±0.44	2.9±0.62	92.11±0.031
0	5.64±1.6	1.73±0.27	2.34±0.36	94.95±0.015

more effective results in depression prevention, treatment, and therapeutics area: (a) establishing future studies on larger and more balanced datasets, (b) recording task-related EEG signals to investigate the effect of different tasks (c) extracting informative features, such as auditory evoked potential and visual evoked potential.

REFERENCES

[1] W. H. Organization, "Depression and other common mental disorders: global health estimates," World Health Organization, Tech. Rep., 2017.

[2] K. Kamenov, F. F. Caballero, M. Miret, M. Leonardi, P. Sainio, B. Tobiasz-Adamczyk, J. M. Haro, S. Chatterji, J. L. Ayuso-Mateos, and M. Cabello, "Which are the most burdensome functioning areas in depression? a cross-national study," *Frontiers in Psychology*, vol. 7, p. 1342, 2016.

[3] P. Cuijpers, E. Karyotaki, E. Weitz, G. Andersson, S. D. Hollon, and A. van Straten, "The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis," *Journal of Affective Disorders*, vol. 159, pp. 118–126, 2014.

[4] G. Ayano, S. Demelash, K. Haile, M. Tulu, D. Assefa, A. Tesfaye, K. Haile, M. Solomon, A. Chaka, L. Tsegay *et al.*, "Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in ethiopia," *Annals of General Psychiatry*, vol. 20, no. 1, pp. 1–10, 2021.

[5] Y.-P. Wang and C. Gorenstein, "Psychometric properties of the beck depression inventory-ii: a comprehensive review," *Brazilian Journal of Psychiatry*, vol. 35, no. 4, pp. 416–431, 2013.

[6] K. L. Smarr and A. L. Keefer, "Measures of depression and depressive symptoms: beck depression inventory-ii (BDI-II), center for epidemiologic studies depression scale (CES-D), geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), and patient health questionnaire-9 (PHQ-9)," *Arthritis Care & Research*, vol. 63, no. S11, pp. S454–S466, 2011.

[7] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of bmd and adhd patients using their EEG signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956–1963, 2011.

[8] A. D. Nazhvani, R. Boostani, S. Afrasiabi, and K. Sadatnezhad, "Classification of adhd and bmd patients using visual evoked potential," *Clinical Neurology and Neurosurgery*, vol. 115, no. 11, pp. 2329–2335, 2013.

[9] D. Jarchi, R. Boostani, M. Taheri, and S. Sanei, "Seizure source localization using a hybrid second order blind identification and extended rival penalized competitive learning algorithm," *Biomedical Signal Processing and Control*, vol. 4, no. 2, pp. 108–117, 2009.

[10] M. R. Islam, M. A. Rahim, H. Akter, R. Kabir, and J. Shin, "Optimal imf selection of emd for sleep disorder diagnosis using EEG signals," *Proceedings of the 3rd International Conference on Applications in Information Technology*, pp. 96–101, 2018.

[11] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," *Applied Sciences*, vol. 9, no. 14, p. 2870, 2019.

[12] O. Al Zoubi, A. Mayeli, A. Tsuchiyagaito, M. Misaki, V. Zotev, H. Refai, M. Paulus, J. Bodurka, R. L. Aupperle, S. S. Khalsa *et al.*, "EEG microstates temporal dynamics differentiate individuals with mood and anxiety disorders from healthy subjects," *Frontiers in Human Neuroscience*, vol. 13, p. 56, 2019.

[13] J. P. Amezquita-Sanchez, N. Mammone, F. C. Morabito, S. Marino, and H. Adeli, "A novel methodology for automated differential diagnosis of mild cognitive impairment and the alzheimer's disease using EEG signals," *Journal of Neuroscience Methods*, vol. 322, pp. 88–95, 2019.

[14] S. D. Puthankattil and P. K. Joseph, "Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy," *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 04, p. 1240019, 2012.

[15] M. Ahmadlou, H. Adeli, and A. Adeli, "Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men," *Clinical EEG and Neuroscience*, vol. 44, no. 3, pp. 175–181, 2013.

[16] "Fractality analysis of frontal brain in major depressive disorder," *International Journal of Psychophysiology*, vol. 85, no. 2, pp. 206–211, 2012.

[17] O. Faust, P. C. A. Ang, S. D. Puthankattil, and P. K. Joseph, "Depression diagnosis support system based on EEG signal entropies," *Journal of Mechanics in Medicine and Biology*, vol. 14, no. 03, p. 1450035, 2014.

[18] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, and A. Adeli, "Computer-aided diagnosis of depression using EEG signals," *European Neurology*, vol. 73, no. 5-6, pp. 329–336, 2015.

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

[20] G. M. Bairy, O. S. Lih, Y. Hagiwara, S. D. Puthankattil, O. Faust, U. Niranjan, and U. R. Acharya, "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 8, pp. 1857–1862, 2017.

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

[22] X. Li, R. La, Y. Wang, J. Niu, S. Zeng, S. Sun, and J. Zhu, "EEG-based mild depression recognition using convolutional neural network," *Medical & Biological Engineering & Computing*, vol. 57, no. 6, pp. 1341–1352, 2019.

[23] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach," *Cognitive Neurodynamics*, vol. 15, no. 2, pp. 239–252, 2021.

[24] J. F. Cavanagh, A. Napolitano, C. Wu, and A. Mueen, "The patient repository for EEG data+ computational tools (pred+ ct)," *Frontiers in Neuroinformatics*, vol. 11, p. 67, 2017.

[25] J. F. Cavanagh, A. W. Bismark, M. J. Frank, and J. J. Allen, "Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed EEG," *Computational Psychiatry*, vol. 3, pp. 1–17, 2019.

[26] M. Tohen, C. Bowden, A. A. Nierenberg, and J. Geddes, *Clinical trial design challenges in mood disorders*. Academic Press, 2015.

[27] Zajac and Paszkiel, "Using brain-computer interface technology as a controller in video games," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 10, no. 3, pp. 26–31, 2020.

[28] X. Ding, X. Yue, R. Zheng, C. Bi, D. Li, and G. Yao, "Classifying major depression patients and healthy controls using EEG, eye tracking and galvanic skin response data," *Journal of Affective Disorders*, vol. 251, pp. 156–161, 2019.

[29] S. Fitzgibbon, D. DeLosAngeles, T. Lewis, D. Powers, T. Grummett, E. Whitham, L. Ward, J. Willoughby, and K. Pope, "Automatic determination of emg-contaminated components and validation of independent component analysis using EEG during pharmacologic paralysis," *Clinical Neurophysiology*, vol. 127, no. 3, pp. 1781–1793, 2016.

[30] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster independent component analysis by preconditioning with hessian approximations," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4040–4049, 2018.

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

[32] D. Jiang, Y.-n. Lu, M. Yu, and W. Yuanyuan, "Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement," *Expert Systems with Applications*, vol. 121, pp. 188–203, 2019.

[33] D. Jiang, Y. Ma, and Y. Wang, "A robust two-stage sleep spindle detection approach using single-channel EEG," *Journal of Neural Engineering*, vol. 18, no. 2, p. 026026, 2021.

[34] P. Pandey and K. Seeja, "Subject-independent emotion detection from EEG signals using deep neural network," *International Conference on Innovative Computing and Communications*, pp. 41–46, 2019.

[35] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in Biology and Medicine*, vol. 106, pp. 71–81, 2019.

[36] Q. Wang, D. Zhao, Y. Wang, and X. Hou, "Ensemble learning algorithm based on multi-parameters for sleep staging," *Medical & Biological Engineering & Computing*, vol. 57, no. 8, pp. 1693–1707, 2019.

[37] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks*, vol. 124, pp. 202–212, 2020.

[38] A. v. d. 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," *arXiv preprint arXiv:1609.03499*, 2016.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[40] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[42] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[43] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil, and U. R. Acharya, "Automated depression detection using deep representation and sequence learning with EEG signals," *Journal of Medical Systems*, vol. 43, no. 7, pp. 1–12, 2019.

[44] L. R. Trambaiolli and C. E. Biazoli, "Resting-state global EEG connectivity predicts depression and anxiety severity," *in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3707–3710, 2020.

[45] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales prediction using linear and KNN regression," *in Advances in Machine Learning and Computational Intelligence*, pp. 321–329, 2021.

[46] S. Kumar, M. Yadava, and P. P. Roy, "Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction," *Information Fusion*, vol. 52, pp. 41–52, 2019.

[47] H. G. Daoud, A. M. Abdelhameed, and M. Bayoumi, "Automatic epileptic seizure detection based on empirical mode decomposition and deep neural network," *in 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 182–186, 2018.

[48] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style convnets great again," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13 733–13 742, 2021.