Journal of Theoretical and Applied Information Technology
31st December 2023. Vol.101. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTIMODAL SIGNAL ANALYSIS FOR DEPRESSION AND ANXIETY PREDICTION: A HYBRID CNN-RNN APPROACH

MAMTA KUMARI[1], DR. MURALI KRISHNA PUTTAGUNTA[2], ANISHA J[3], DR. TAVITI NAIDU GONGADA[4], DILFUZA GULYAMOVA[5]

Assistant professor of CSE(AI-ML),Panipat institute of engineering and technology (PIET), Samalakha[1]
Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, INDIA[2]
MBA, Assistant Professor, Panimalar Engineering College, Chennai[3]
Assistant Professor, Dept of operations, GITAM school of Business, GITAM (Deemed to be) University, Visakhapatnam, India[4]
Computer Engineering Department, University of Information Technologies in Tashkent, Tashkent, Uzbekistan[5]

Email- mamta1.tarar@gmail.com[1], murali93940@gmail.com[2], mba.anisha03@gmail.com[3], tgongada2@gitam.edu[4], gulyamova2710@gmail.com[5]

ABSTRACT

Anxiety is a serious global problem. The cornerstones of management are drug and psychological therapy. Because some individuals will respond to traditional medications, regardless of their efficacy, there is a demand for alternative strategies for preventing and treating depression. The physiological markers that can be employed in combination with a method to recognize depression are described in this study. This article contributes to the successful instructional approach by optimizing a convolutional neural network model. In this instance, the attributes are extracted using a Convolutional Neural Network (CNN), which increases the classification accuracy and predictability of sadness. RNN forecasts the existence or absence of depression. This unique approach aims to reduce training time and improve accuracy by combining the use of convolutional neural networks with bi-directional short-term and long-term memory. Using the suggested method, the CNN-RNN classifier uses ECG data to predict sadness and anxiety. The proposed method for identifying depressive illnesses effectively obtained 99.6% accuracy, 99.5% recall 99% precision, and a 99.7%F1 score using a combination of evaluation of heterogeneous emotional signals to anticipate melancholy and assess stress levels. The success of the proposed strategy is confirmed through the comparison of its outcome indicators with the validation data.
Keywords: *Anxiety; Psychological Treatments; CNN; RNN; Melancholy*

1. INTRODUCTION

Depression is an ingrained mental illness. It is a severe long-term wellness disease, distinct from typical erratic emotions and fleeting feelings to everyday occurrences. Constant irritability is frequently seen. Depression, a lack intriguing in and happiness in things, a low sense of individuality, and a refusal to handle daily obligations [1]. Many times, the body's response to unexpected circumstances becomes anxious or stressed, which is a normal physiological reaction. Anxiety can, however, last a lifetime if there are aberrant responses to commonplace stressors. The term "stress-related diseases" refers to a group of aberrant responses that comprises phobias, social anxiety, panic attacks, and generalized disorders of anxiety. In their lifetime, greater than 30% of people have experienced an anxiety problem. Numerous factors, including stressful situations from the past, weariness, burnout from a job, loneliness, abuse, etc., can contribute to anxiety, which can lower quality of life and decrease work productivity [2].

One of the primary manifestations of depressive disorders, anhedonia, is the inability to enjoy life. This symptom frequently recurs after a period of remission. It looks crucial for resilience against recurrence to be able to experience rewards. Positive emotional reactions lessen the sensitivity to stress and indicate a successful outcome from antidepressant therapy. Additionally, pleasure serves a crucial motivating purpose by reinforcing behaviors that may trigger events. Patients with MDD frequently report lacking desire to achieve their goals or having trouble finding pleasure in ordinarily enjoyable activities, such as consummate anhedonia. In addition, people with MDD have a hard time picking up new habits that could keep them healthy or make them feel better [3]. Applying surveys for self-reporting and conducting clinical discussions are the conventional methods to identify anxiety and stressful situations. This approach evaluated 374 college-level pupils using a 21-question variant of the Depressive Distress Pressure Scale. The findings indicate that kids are most stressed out by anxiousness, performance in school, and a desire to achieve their goals. According to the study's findings, anxiety-related issues frequently interfere with educational institutions learner's capacity to succeed. Organizations must consequently constantly evaluate the emotional well-being of their students to apply protection strategies that will enhance their academic achievement [4].

Psychological as well as bodily symptoms of depression or anxiety are typically discernible. The physiological responses include the response of the sympathetic nervous system, the glands that produce adrenaline, and the production of hormonal substances, whereas psychological manifestations include irritation, nervousness, disruption, and disorientation. Common signs of anxiety-related disorders, which are a subset of mental illnesses, include worry, avoiding social situations, panic attacks, fear, distress, and insomnia. Employing a difference of psychological cues, which are mainly dynamic video and audio triggers and static, can help decipher these signs and symptoms [5].

Experimental including translational study evidence indicates that depressive disorders and stress are related to boosted immune system function, improved leukocyte function, and the release of pro-inflammatory cytokines. To identify stress as well as categorize emotions, scientists have come up with a variety of techniques for assessing physiological indicators captured by the sensors linked to human bodies [6]. 43 million further years of physical activity might be achieved, according to an up-to-date review of information from 36 nations, if the cost of prescriptions for depression and other psychiatric disorder conditions is increased to US$147 billion from 2016 through 2030. A US$310 billion estimate has been made for the economic repercussions of this period. A US$230 billion increase in gross domestic product might occur from feeling depressed therapeutic expenses being pushed up [7].

While a particular individual exhibits the majority of the depressive symptoms such as anhedonia, overbearing guilt, thoughts of committing suicide, disturbances in slumber and appetite, psychological intellectual disability, poor concentration, and exhaustion for at least two weeks, they are said to have a major depressive illness (MDD), according to the Diagnostic and Statistical Manual of Mental Disorders-5. For an official diagnosis of MDD, one of these criteria—a depressed psychological state or anhedonia—must be met (or a combination) [8]. Physical manifestations of feelings include the process of respiration the Galvanic Skin Response (GSR), temperature, electrocardiogram (ECG), electroencephalogram (EEG), and so on. When an identical emotion is presented with the same modality indication, different participants may respond physiologically differently. Moreover, different physiological indicators have different weights concerning different moods. Better outcomes could be achieved by combining and complementing multiple modalities of data utilizing fusion algorithms [9].

For measuring and identifying aberrant heart rhythms, an ECG is the most effective technique. The repetitive and unpredictable ECG signal has a brief lifetime. It is the most crucial harmless physiological indicator for diagnosing and monitoring CVD. Diagnosing aberrant cardiac rhythms involves identifying two disorders. when the connective tissue that conducts electrocardiograms is harmed first. Second, if an electrolyte imbalance results in a change in the circulatory rhythm. It can be used to pinpoint the location of the myocardial infarction while evaluating the condition. The categorization findings from this work demonstrate the potential of the proposed CNN-RNN model for determining depression and anxiety levels in ECG pictures. In the meantime, one of the markers for melancholy detection can be the categorization findings of normal and abnormal ECGs. ECG signal processing's main objective is to fundamentally eliminate disturbance to produce a precise diagnosis. Heart rate variability (HRV), morphological characteristics, frequency of time domain feature analysis, and wavelet analysis have all been used for obtaining and analyzing ECG data in a large number of studies [10].

For 2D image analysis tasks including acknowledgment, categorization, and forecasting, CNN is the deep-learning model type that performs the best. High-level characteristics are extracted by CNN using numerous hidden convolution layers after it accepts 2D data as the source. Some research

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

investigations have transformed signals of 1D into 2D data to distribute physiological signals through a model of CNN. With the use of huge data, deep learning techniques based on CNN might yield good results. It is not necessary to extract/select features, reduce features, or classify features for CNN. The neural network performs discrimination and feature learning in a supervised way. The classification of physiological data is another area where deep learning approaches are gaining recognition and fascination from researchers [11].

In particular, when using numerous layers for every step in the process, RNNs, those with long short-term memory (LSTM) units, are effective at analyzing time series applications. The key contribution of the described outline is given as follows:

- The study focuses on utilizing multimodal psychological signals, such as ECG records, to identify depression.
- The research introduces a hybrid CNN-RNN model, optimizing the learning process and improving the accuracy of depression prediction.
- The CNN model is utilized for feature extraction, enhancing classification accuracy, and the ability to predict depression.
- By incorporating RNN in conjunction with the CNN, the proposed approach aims to enhance the accuracy of depression prediction

The configuration of the given essay is as follows: Section 2 contains the related works that are framed to understand the proposed paper with the existing methods while Section 3 elaborates on the problem statement. Section 4 depicts the proposed CNN-RNN architectures. The results and performance metrics are tabulated and graphically represented in section 5. Finally, in chapter 6, the conclusion is presented.

2. RELATED WORKS

Saeed et al proposed psychological labeling is offered as a basis for an EEG-based classification of long-term stress. Human stress or depression can be categorized into a supervised machine learning model created here. This strategy involved a total of 33 volunteers. Using an EMOITV Insight headset and closed eyes for three minutes, the resting state EEG data for each subject were recorded. Participants completed the PSS-10 questionnaire after the EEG signal was recorded, and an expert in psychology then interviewed them. An interview typically takes 25 minutes. Each participant was assembled into either the control group or the stress group by the psychology expert based on their PSS scores and interviews. The pre-processing procedure removed all noise from the obtained EEG data. From the signals at each electrode, neurophysiological characteristics such as alpha (α), low beta (βl), beta (β), gamma (γ), delta (δ), theta (θ), and relative gamma (RG) power were retrieved. These features were used to calculate the frontal, alpha, temporal, and beta asymmetries as well as the alpha asymmetry. The classification of human stress was carried out using five algorithms of supervised machine learning: KNN, NB, SVM, MLP, and LR. The felt stress scale and expert rating, using both interview and PSS results simultaneously, were two different labeling techniques. The classification accuracy was increased by up to 85.20% using the expert evaluation-based labeling method. The drawback of this approach is that they can't accurately assess neuronal activity below the brain's top layers [12].

Salkevicius et al developed different levels of anxiety detected using hypothesized physiological cues and a treatment system that simulates actual life. The potential applications of this type of VRET technology were emphasized by combining VRET with the detection of physiological strain. The cloud-based VRET system that has been built includes a framework for recognizing anxiety levels. The technique involved collecting the physiological signs of 30 people. The gathered data were used to train a four-level anxiety recognition model; "mild," "moderate," "low," and "high" may relate to the levels of anxiety rather than different classes of anxiety illness. Here, a support vector machine classifier using signal fusion was able to attain cross-subject accuracy of 80.1% and 86.3%. The integration of the framework for predicting anxiety levels into the user interface is not the primary goal of this approach [2].

Song et al. [13] elaborate on the pressing issue of developing objective and readily available diagnostic tools for depression analysis. It highlights the shortcomings of traditional clinical diagnosis methods, which are subjective, complicated, and rely heavily on clinician involvement. The study introduces a novel approach that overcomes existing barriers in automatic depression analysis algorithms. By extracting multi-scale video-level features based on automatically detected human behavior primitives, the proposed spectral representations, including spectral heatmaps and spectral vectors, capture the temporal dynamics of expressive behavior. These spectral representations are then fed to Convolution Neural Networks (CNNs) and Artificial Neural Networks (ANNs) for depression analysis. The experiments conducted on benchmark datasets demonstrate not only state-of-the-art accuracy in estimating depression severity but also the importance of the user's interview task in influencing the analysis. The fusion of multiple tasks leads to the highest accuracy, while longer tasks are shown to be more informative than shorter tasks up to a certain point. this approach presents a promising and innovative approach to automatic depression analysis, paving the way for potential advancements in mental health diagnostic tools with enhanced objectivity and efficacy.

Jan et al. [14] present an innovative approach for monitoring depression using artificial intelligence. The system aims to predict Beck depression inventory II (BDI-II) scales based on visual and vocal expressions. The research leverages deep learning methods to extract key visual features from facial expression images and utilizes spectral low-level descriptors and mel-frequency cepstral coefficients to capture vocal expressions from short audio segments. Additionally, a novel feature dynamic history histogram (FDHH) is introduced to capture temporal movement in the feature space. The fusion of FDHH and audio features through regression techniques enables the prediction of BDI-II scales. The proposed method is tested on the Audio/Visual Emotion Challenges 2014 dataset, tailored to focus on the study of depression. The results demonstrate superior performance compared to existing methods on the same dataset. Applying artificial intelligence to analyze both visual and vocal expressions, the system provides a promising tool for automatic depression level analysis. The study showcases the potential of AI in understanding and predicting mental health states through human gestures and facial expressions. the paper contributes to the advancement of AI-based approaches for mental health assessment and signifies the importance of leveraging multimodal data for more accurate depression monitoring.

Zitouni et al suggested automatic recognition of severe depression disorder from physiological signs in the heart and lungs. This method uses relevant multi-modal physiological signals, such as the electrocardiogram (ECG), finger photoplethysmography (PPG), as well as respiratory signals (RSP), to automatically determine the disorder severity in MDD patients. This research included data obtained from 88 participants. Each subject's multi-modal physiological signals, such as PPG, ECG, and RSP signals, were collected before being pre-processed. Eleven nonlinear characteristics were identified after the signals underwent Discrete Wavelet Transform (DWT) that was carried out after they were decomposed up to six levels. Marginal Fisher Analysis was used to reduce the feature set after the features were ranked using the analysis of variance test, and the remaining reduced features were then ranked once more to determine which features were the most discriminatory. Using a 10-fold cross-validation approach, the classifiers' performance was assessed. This research represents a step toward the use of automated technologies in the diagnosis and follow-up of MDD patients in a tailored, wearable healthcare environment. Three classifications were made with an accuracy of up to 85.2%. The disadvantage is that it doesn't allow for the use of deep learning on less lengthy pieces of information to additionally recognize individuals with MDD but also categorize as well as analyze the extent of depression along with suicidal thoughts determined by both of the gained psychiatric as well as self-reported information ratings of depressive symptoms, for ongoing monitoring and assistive diagnostics [15].

Trotzek, Koitka, and Friedrich [16] present a crucial and timely topic, as depression remains a significant global health concern. The study explores the potential of machine learning models to detect early indications of depression in social media messages, recognizing the importance of language usage in expressing mental health issues. The paper compares the performance of a convolutional neural network (CNN) using different word embeddings with a classification approach based on user-level linguistic metadata. The ensemble of both methods is demonstrated to achieve state-of-the-art results in early detection. This paper critically evaluates the ERDE score, a popular metric for early detection systems, revealing its limitations in shared tasks. A modified metric is proposed, contributing to more accurate evaluations of early detection performance. the research introduces a new word embedding trained on a large corpus of the same domain, enhancing the model's performance. the paper presents a comprehensive and well-executed investigation into early depression detection using neural networks and linguistic metadata. The findings contribute significantly to advancing the field of mental health analysis in text sequences and have implications for

Journal of Theoretical and Applied Information Technology
31st December 2023. Vol.101. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

creating effective and timely interventions for individuals at risk of depression.

Rodriguez-Arez et al proposed the direction of a physiologically oriented method of identifying stress and anxiousness in educational circumstances. With the help of a board for Arduino and inexpensive sensors, this research will pinpoint the physiological traits that can be utilized to anticipate stressful situations including anxiety levels in learning scenarios. 21 students were primarily utilized for testing, which resulted in the development of a causing stress procedure including the analysis of 21 physiological aspects involving five signals. To further gauge each student's level of anxiety, the State-Trait Anxiety Inventory (STAI) was employed. The physiological feature subset with the best accuracy in identifying stress and anxiety phases was determined by comparing four algorithms. Since this was going on, the SVM classifier, utilizing data obtained from three physiological features and also the reaction of the galvanic skin (GSR) signal was able to identify anxiousness with a success rate of more than 95%. While just a single kind of anxiety activity was evaluated, this research's design of experiments may have certain limitations. However, this investigation decided to utilize a different kind of academic task, to run hypothetical models, and to gauge the student's anxiety levels about the difficulty of the task and their academic achievement. The application of computerized feature selection techniques is also included in this approach [17].

2.1. Problem Statement

The probability of premature mortality is regularly proven to be increased by psychological disorders, of which melancholy is a major contributor [18]. Individuals who are depressed may have recurring thoughts of suicide, sentiments of inadequacy or frequent guilt, a lack of enthusiasm for and enjoyment from daily activities, considerable weight loss or gain, excessive sleeping, exhaustion, insomnia, and/or a diminished capacity to focus on anything [19]. The drawback of the existing approach is that they can't accurately assess neuronal activity below the heart's rate recordings it may be due to nervousness [12]. To solve this issue, a novel combined Convolutional RNN mechanism is proposed, which can extract all the spatio-temporal features from the ECG recordings. A detailed explanation of the proposed algorithm is given below.

2.2. Research Questions

1. How effective is the hybrid CNN-RNN model, incorporating multimodal psychological signals, particularly ECG records, in accurately identifying and predicting depression compared to traditional models or single-modality approaches?

2. To what extent does the use of a CNN model for feature extraction contribute to improving the accuracy of depression classification when applied to multimodal psychological signals?

3. How does the integration of RNN with CNN impact the predictive performance and overall accuracy in identifying depression?

3. PROPOSED CNN-RNN MODEL-BASED DEPRESSION PREDICTION

Initially, the physiological signals like the ECG dataset are preprocessed through the Butterworth low pass filter and then the feature is extracted by using a Convolutional Neutral Network and then classification as well as prediction is employed by the CNN-RNN model. Figure 1 shows the block diagram of depression and anxiety detection through a hybrid CNN-RNN model

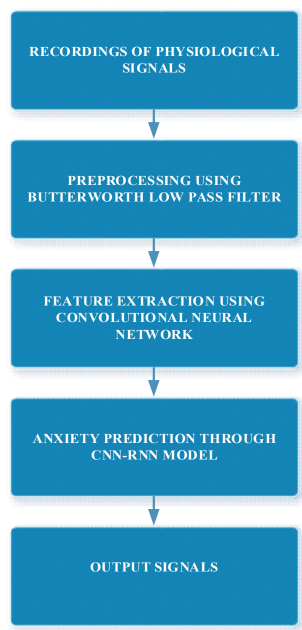

Figure 1: Block Diagram of depression and anxiety detection through hybrid CNN-RNN model

3.1. Data Collection

At the National Institute of Electronic Science and Technology of Chinese in Chengdu, Sichuan Province, China, 20 researchers of males as a whole, free from psychological and cardiovascular conditions, were shortlisted throughout the intended research project. The following average body mass index (BMI) varied from 18.289 to 29.068 Kilograms/m2, while they had no previous record of taking medicine or having problems with sleeping. The ages of the participants varied from 23 to 35, as well as they were 162 to 187 cm tall, 48 to 88 kg, in addition to BMI reached from 48 to 88 kg. For 12 hours before the ECG record keeping, each participant was instructed never to consume any alcohol, coffee, or tea. The research proposal was authorized by the university ethical committee after receiving informed permission from any participants in addition to the protocols for the study. The Chalder Fatigue Score (CFS), Depressive disorders, Specific Fatigue Scale (SFS) anxiety, and Stress Scale (DASS) were completed by each of the participants earlier than the experiment to gauge their level of psychological strain and other mental health problems. The Cold Pressor Stress (CPS) test, Montreal Imaging Stress Task, Stress Evaluation Test, and DASS have all been used in the past to identify mental stress. In the experiment, we captured heart rhythms using an electronic T-shirt. Researchers decide on a 2:1 ratio between mental stress and regular conditions. Immediately following 12 hours of nonstop work in the lab, we recorded the circumstances of mental fatigue. In addition, we documented the typical morning circumstances when we weren't in the lab. At one-day intervals, we monitored people's mental exhaustion[20].

3.2. Data Preprocessing

A significant amount of noise can often be found in the ECG signals that the device collects. Power frequency interference, baseline drift, and interference with ECG are examples of common noise. As a result, before categorization, this signal must be filtered. Various paradigms may need to be aligned, normalized, and filtered during this process. A low-pass filter was employed to eliminate ECG interference, which is a notch filter to eliminate power frequency interference, and a median filter to eliminate baseline drift after evaluating several filtering techniques in the experiment. Following the application of a notch filter to remove noise from 50 Hz power frequencies, a median filter with a window size of 109 was used to remove baseline drift. The first step in the process was to employ a Butterworth low-pass filter to remove noise at frequencies above 80 Hz. All ECG signals have been resampled to the same 360 Hz rate of sampling to remove offset and standardize this signal's amplitude. To remove the effect of time latency in the filtration procedure, 50 sampling points were eliminated from both before and following the ECG signals [21].

3.3. Feature Extraction

Convolutional neural networks (CNNs) have several different kinds of levels: accumulating, convolution, and completely integrated. Three fully connected layers, five pooling layers, and five convolutional layers make up the CNN algorithm used in this study. During learning, the filter's weights are changed. It is a weighted vector used to convolve the input. Convolution and pooling processes each have a filter set at a value of 5, correspondingly. As a result of their ability to produce the most accurate results, filters of 5 and 2 were selected. In addition, the stride (the quantity by which the sampling point window is adjusted in each operation) is fixed at 1 for the convolution operation and at 2 for the pooling operation, which will be discussed in more detail later. As the model's accuracy is improved, the parameters are discovered through trial and error [22].

1) Convolutional layer:

By swiping the kernel over the input during this layer, convolution is carried out to produce a convolved output employing the calculation below in eqn. (1).

$$a_n = \sum_{n=0}^{N-1} f_m k_{n-m} \tag{1}$$

The signal, output, filter, and the total number of data points present in the input, k, are denoted by the letters k, f, a, and N, correspondingly. The nth output element that is being calculated is indicated by the subscript m, while the nth component of the filter vectors is indicated by m. The sample f_m is multiplied by the sample from the input signalk_{n-m}while n can vary from 0 to N-1. The output is created by adding together each of these items. To extract essential characteristics from the starting point EEG signal for the technique's learning, the convolutional neural layer's goal is to evaluate the signal. Following this process, a leaking rectifier unit is used. The activation functions for layers 1, 3, 5, 7, 9, 11, and 12 use it. Leaky Elu's primary purpose is to sustain the durability of the calculation process by making it more sensitive to the noise that exists in the EEG signals. The method of activating the function for Layer 13 is the SoftMax functioning.

2) Pooling layer:

The important characteristics are still present although the overall dimension of the map of features is reduced thanks to this layer. In the present research, a max-pooling operation is used, which implies that following each maximum pooling function, the attribute map's stride 2 window's biggest result is kept.

3) *Fully-connected layer:*

Employing the formula below, this layer of neurons links all neurons in that layer to each synapse in the stratum above it.

$$y_i = \sum_j w_{ji} z_j + b_i \qquad (2)$$

The outputs from the layers are denoted by z and y, respectively, wherein w variable b stands for the bias and weight values. To estimate the type of signal by determining the class probability of each ECG signal being classified as regular or depression, the SoftMax algorithm is fed its results from the last completely connected layer. The present research uses a 13-layer CNN model with summarized information on the steps, filtering dimensions, and the number of neurons in each layer. To create the first layer (output neurons of 1996 x 5) from the layer of inputs, a method known as convolution is used having a 5-filter size (stride 1). The second level's synapses are then reduced to 998 times five by a maximum pooling procedure that is carried through the initial level. The next one (layer 3) follows by running another cycle of combing. Following convolution, which is layer 4 with 497 x 5 neurons is obtained by once more applying the maximum pooling process. To create layers 5, 6, 7, 8, 9, and 10, three further convolutions then max-pooling procedures are carried out alternatively. 80 neurons in level 11, the top completely linked layer, are linked to layer 10 via layer 10, while layer 12 is connected via layer 11 via layer 40. Two neurons that produce signals from layer 12 that signify healthy and depressed have connections to the layer that follows it last.

3.4. Model Prediction

As time-sequential expressions for CNN in the extraction process, physiological procedure data can be expressed in image datasets. Machine learning approaches can encrypt time-sequential data pretty effectively. We employed RNN because it is the most effective technique for dealing with data that is sequential. RNN has constant connections to undetected future, present, and past variables. Gradient disappearing problems might happen in retention, which is crucial in neural network design. LSTM can be used to fix problems with storage management. The RNN's fundamental structure is composed of 60 LSTM algorithm implementations [23]. Fig. 2 illustrates the basic layout of an RNN that has Sixty of the LSTM algorithm components as well. Recurrent Neural Network is referred to as RNN. For processing sequential data, such as time series data or natural language, it uses a particular type of neural network architecture. RNNs have connections that enable information to flow in loops, as opposed to feed-forward neural networks that process data in a one-directional flow. This allows RNNs to maintain an internal memory or context of prior inputs. The ability of RNNs to recognize and make use of temporal dependencies in the data is their distinguishing characteristic. Each RNN unit maintains an internal state or hidden state in addition to producing an output from an input. The network can take into account the context when making predictions or decisions because the hidden state serves as a memory that stores data about prior inputs. RNNs can handle variable-length sequences because of their recurrent nature, which allows them to process inputs of various lengths in stages. RNNs are thus well suited for tasks like text generation, sentiment analysis, machine translation, and speech recognition.

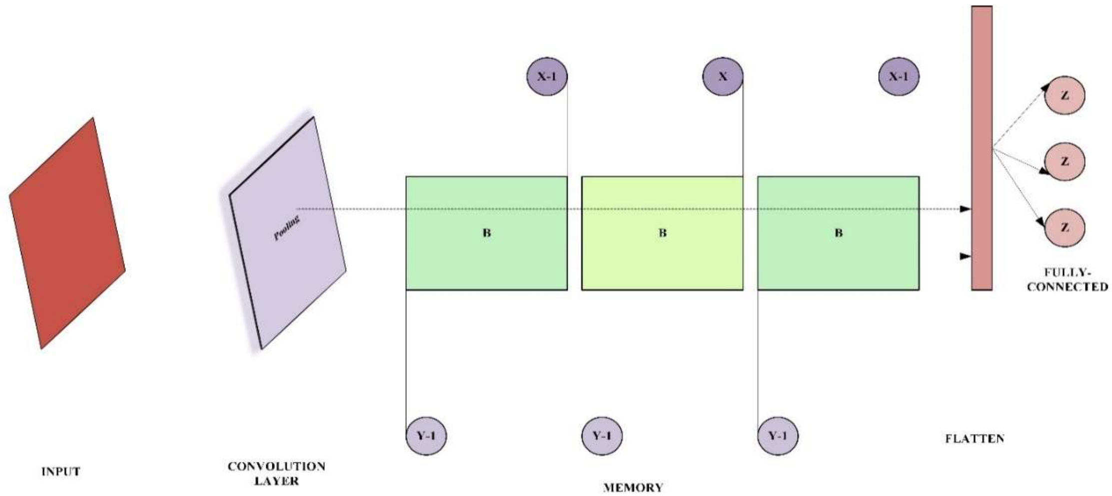

Figure 2: Architecture of CNN-RNN Model

There are two states: a cell condition as well as the gate phase in each LSTM block and as shown in Fig. Input, Forget, and Result gateways make up the three types of gates that make up the gate condition. The ability of a network using LSTM to retain and comprehend past data nevertheless does not translate into the ability to use new information to aid in making final recommendations. It is therefore necessary to make additional changes to LSTM networks with two-way communication by giving the network the ability to perform both backward and forward computation [24]. Eqn. (3) serves to illustrate how the input gate is implemented.

$$I_t = \sigma(A_{LI}L_t + V_{HL}H_{t-1} + c_t\ ) \quad (3)$$

Here, A is the input from the user, V is the weighted matrix, c is the bias function, t is a date and time stamp, and the logistic function is performed using. As seen in the illustration of the input gateway Eqn. (4)

$$F_t = \sigma(A_{LF}L_t + V_{HF}H_{t-1} + c_f\ ) \quad (4)$$

A cell state vector X that can be expressed in a manner that represents the LSTM in Eqn. (5)

$$X_t = F_tZ_{t-1} + F_t panh(A_{Lz}L_t + V_{Hz}H_{t-1} + c_z\ ) \quad (5)$$

As seen in the formula, the gate for output O is capable of being described in Eqn. (6)

$$O_t = \sigma(A_{LO}L_t + V_{HO}H_{t-1} + c_o\ ) \quad (6)$$

The concealed stratum can be depicted in the following way in Equation (7)

$$h_t = O_t panh(Z_t) \quad (7)$$

We implemented a concentrated barrier on the LSTM components preceding attaching the thicker layer, as depicted below in Equation (8)

$$a(attention)_t = LSTM(h_t, a(attention)_{t-1}) \quad (8)$$

The justification for implementing the focused attention methodology is that the topmost layer of the LSTM algorithm is utilized to increase the model's accuracy through the attentiveness method. The focused learning strategy prioritizes beneficial data over significant data in the current goal-oriented. The output outcomes stated in Equation (9), wherein "W" is the weighted function while "c" is the bias function, are capable of being calculated using a "SoftMax" algorithm.

$$Output = SoftMax(W_RB_R + c_R) \quad (9)$$

4. RESULT AND DISCUSSION

With the assistance of mathematical words for quantifying what was discovered, this research offers a thorough description of the results of experimentation generated by the suggested methodology. Using the ECG records, the verification process was run. 20 male scientists made up the entire dataset, which was free of cardiac and psychological diseases. Utilizing a wireless T-shirt, this technique recorded heartbeats. A split of two to one between mental stress and usual circumstances is chosen by scientists. We observed conditions of mental exhaustion immediately after twelve hours of uninterrupted lab activity. Three classification techniques—LSTM, SVM, and DCNN+LSTM—were evaluated against the system that was proposed to see how well it performed. Four indicators, including accuracy, F1 score, precision, and recall, are used to evaluate how well CNN and RNN predict its interpretation of physiological information.

4.1. Evaluation of performance metrics

Four criteria for analysis were used in the study to rate the simulations. The recommended approach establishes the precision, recall, accuracy, and F-1 measurement to improve the effectiveness of our prediction model. Following is a full explanation of the test criteria. Each technique can perform differentially in terms of the correctly classified incidents, as demonstrated in eqn. (10), which measures reliability as the number of correctly anticipated examples divided by the total amount of forecasts produced by the algorithm.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Recall estimates the percentage that has positive those suffering from depression out of every positive instance for which the simulation was successful in making a precise prediction. Multiplied by the TP by the combined amount of the TP in addition to FN, as well as its eqn. (11), which yields a recall.

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

Correct forecasts show anticipated positive results. Precision represents the total amount of persons who have been classified as having depression, including people who were anticipated to have the condition and it is given in eqn. (12)

$$Precision = \frac{TP}{TP+F} \quad (12)$$

Due to the reality that they analyze different properties that are reflected in Eqn. (13), which is the precision and recall of the F-Measure value allows it to assess the harmonization of two components.

$$F1score = \frac{2*Recall*Precision}{Recall+precisio} \quad (13)$$

Out of all the truly beneficial results, the number of results that were precisely identified as positive is referred to as TP in the formulas above. TN out of all the genuinely negative data, as many were accurately labeled as negative The number of values that the simulation incorrectly identified as being negative despite the fact they were positive in the dataset is called false negatives, or FN. The number of variables that the approach incorrectly categorized as positive even though they were negatives in the actual data set is called the false positive rate (FP) [25].

Table 1. Experimental result analysis for different parameters with other metrics

Method	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
LSTM [23]	98.9	99	98	98
SVM [26]	90	91	89	93
DCNN+LSTM [27]	98.32	98.34	99.78	97.65
Proposed CNN-RNN	99.66	99.5	99	99.7

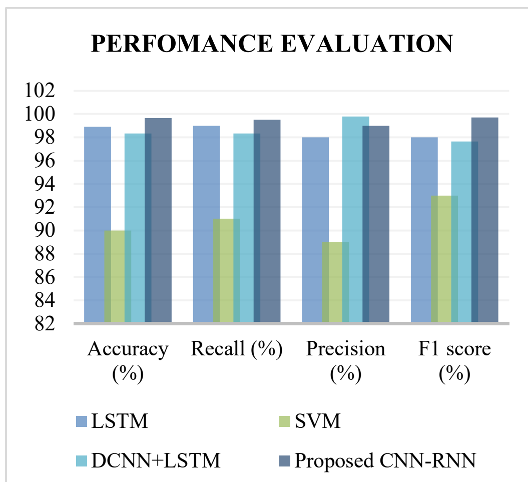

Figure 3: Performance evaluation of various methods of classification

The suggested method's evaluation of efficiency is listed in Table 1. In contrast to existing classification algorithms, the suggested CNN-RNN exhibits higher precision. The suggested approach outperforms previous methods in terms of mean precision and recall. The F1-score is a meaningful classification metric that is determined by the mean value of precision and recall. Figure 3 illustrates the data analysis graphically.

Table 2. Compared results in terms of sensitivity and specificity

Method	Specificity (%)	Sensitivity (%)
LSTM	97.7	99
SVM	86.36	91
DCNN+LSTM	94	98.34

Proposed CNN-RNN	98	99.5

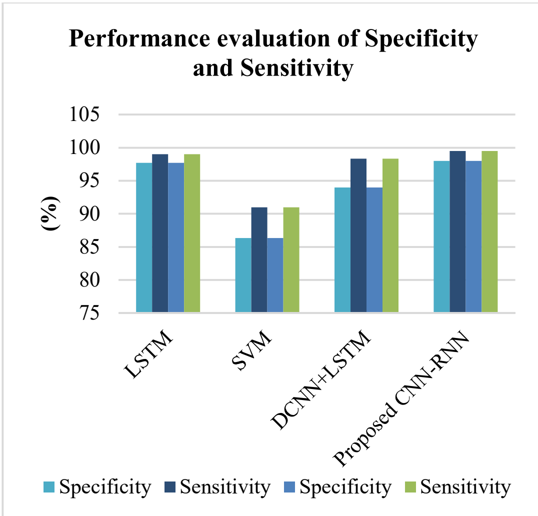

Figure 4: Comparison of Sensitivity and specificity for existing and proposed methods

Table 2. compares the effectiveness outcomes of the suggested approach to those of the current techniques. A graphic illustration of sensitivity and specificity comparison is shown in Graph 2 for easy comprehension**.**

TABLE 3. COMPARED RESULTS IN TERMS OF RMSE

Model	RMSE
Audio-DCNN	5.744 [28]
Audio -DCNN-DNN	5.669 [28]
Video (HDR) -DCNN	5.325 [28]
Proposed model	4.826

Table 3 presents a comparison of the Root Mean Square Error (RMSE) values for different models used in predicting depression and assessing stress levels from multimodal psychological signals. The models evaluated include Audio-DCNN, Audio-DCNN-DNN, Video (HDR)-DCNN, and the Proposed Model.

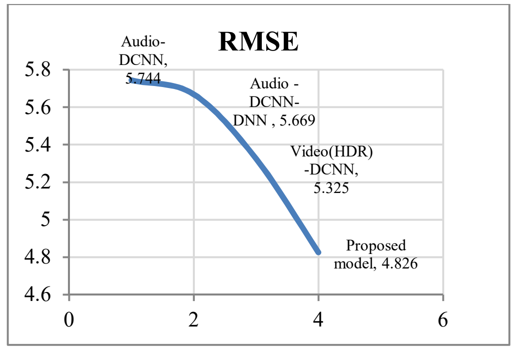

Figure 5: Error Rate Comparison

5. CONCLUSION

The recommended approach is an interesting area for additional research because it is more efficient than existing classifications that offer a higher level of accuracy. The offered paradigm outperforms the previous model by having an accuracy of 99.66%, recall of 99.5%, precision of 99%, with F1-measure of 99.7%. The accepted classifier in the suggested model outperforms the LSTM, SVM, and DCNN+ LSTM techniques. This classification method used an RNN model that was CNN-based. The proposed approach substantially advances the science of depression categorization and forecasting as a whole. Tests on data sets from various sources and industries will next be undertaken to evaluate the suggested strategy. The CNN-based RNN accuracy for prediction can be improved by experimenting with different configurations and hyperparameters. In the future, the research should prioritize cross-dataset validation to assess the model's applicability across diverse sources. Moreover, a detailed exploration of hyperparameter settings and the incorporation of systematic optimization would enhance the model's stability.

REFERENCES

[1] X. Zhu, T. Gedeon, S. Caldwell, and R. Jones, "Detecting emotional reactions to videos of depression," in *2019 IEEE 23rd International Conference on Intelligent Engineering Systems (INES)*, Apr. 2019, pp. 000147–000152. doi: 10.1109/INES46365.2019.9109519.

[2] J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, "Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals," *Electronics*, vol. 8, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/electronics8091039.

[3] H. Geugies *et al.*, "Impaired reward-related learning signals in remitted unmedicated

patients with recurrent depression," *Brain*, vol. 142, no. 8, pp. 2510–2522, Aug. 2019, doi: 10.1093/brain/awz167.

[4] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, p. 105408, Jul. 2020, doi: 10.1016/j.cmpb.2020.105408.

[5] J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, "Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals," *Electronics*, vol. 8, no. 9, p. 1039, Sep. 2019, doi: 10.3390/electronics8091039.

[6] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. S11, p. 285, Dec. 2020, doi: 10.1186/s12911-020-01299-4.

[7] S. Bhatt, A. N. Nagappa, and C. R. Patil, "Role of oxidative stress in depression," *Drug Discov. Today*, vol. 25, no. 7, pp. 1270–1276, Jul. 2020, doi: 10.1016/j.drudis.2020.05.001.

[8] S. J. Yong, T. Tong, J. Chew, and W. L. Lim, "Antidepressive Mechanisms of Probiotics and Their Therapeutic Potential," *Front. Neurosci.*, vol. 13, p. 1361, Jan. 2020, doi: 10.3389/fnins.2019.01361.

[9] S. Zhao, A. Gholaminejad, G. Ding, Y. Gao, J. Han, and K. Keutzer, "Personalized Emotion Recognition by Personality-Aware High-Order Learning of Physiological Signals," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 15, no. 1s, pp. 1–18, Jan. 2019, doi: 10.1145/3233184.

[10] L.-R. Yeh *et al.*, "Integrating ECG Monitoring and Classification via IoT and Deep Neural Networks," *Biosensors*, vol. 11, no. 6, Art. no. 6, Jun. 2021, doi: 10.3390/bios11060188.

[11] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020, doi: 10.1109/TCYB.2019.2905157.

[12] S. M. U. Saeed, S. M. Anwar, H. Khalid, M. Majid, and U. Bagci, "EEG Based Classification of Long-Term Stress Using Psychological Labeling," *Sensors*, vol. 20, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/s20071886.

[13] S. Song, S. Jaiswal, L. Shen, and M. Valstar, "Spectral Representation of Behaviour Primitives for Depression Analysis," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 829–844, Apr. 2022, doi: 10.1109/TAFFC.2020.2970712.

[14] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions," *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 3, pp. 668–680, Sep. 2018, doi: 10.1109/TCDS.2017.2721552.

[15] M. S. Zitouni, S. Lih Oh, J. Vicnesh, A. Khandoker, and U. R. Acharya, "Automated recognition of major depressive disorder from cardiovascular and respiratory physiological signals," *Front. Psychiatry*, vol. 13, 2022, Accessed: Jun. 05, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.970993

[16] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 588–601, Mar. 2020, doi: 10.1109/TKDE.2018.2885515.

[17] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, p. 105408, Jul. 2020, doi: 10.1016/j.cmpb.2020.105408.

[18] S. Dabhane and P. Chawan, "Depression Detection on Social Media using Machine Learning Techniques: A Survey," vol. 07, pp. 2395–0056, Feb. 2021.

[19] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gómez, "A text classification framework for simple and effective early depression detection over social media streams," *Expert Syst. Appl.*, vol. 133, pp. 182–197, Nov. 2019, doi: 10.1016/j.eswa.2019.05.023.

[20] M. B. Bin Heyat *et al.*, "Wearable Flexible Electronics Based Cardiac Electrode for Researcher Mental Stress Detection System Using Machine Learning Models on Single Lead Electrocardiogram Signal," *Biosensors*, vol. 12, no. 6, p. 427, Jun. 2022, doi: 10.3390/bios12060427.

[21] X. Zang, B. Li, L. Zhao, D. Yan, and L. Yang, "End-to-End Depression Recognition Based on a One-Dimensional Convolution Neural Network Model Using Two-Lead ECG Signal," *J. Med. Biol. Eng.*, vol. 42, no. 2, pp. 225–233, Apr. 2022, doi: 10.1007/s40846-022-00687-7.

[22] 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," *Comput. Methods Programs Biomed.*, vol. 161, pp. 103–113, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.012.

[23] A. Amanat *et al.*, "Deep Learning for Depression Detection from Textual Data," *Electronics*, vol. 11, no. 5, p. 676, Feb. 2022, doi: 10.3390/electronics11050676.

[24] Y. Liang, S. Yin, Q. Tang, Z. Zheng, M. Elgendi, and Z. Chen, "Deep Learning Algorithm Classifies Heartbeat Events Based on Electrocardiogram Signals," *Front. Physiol.*, vol. 11, p. 569050, Oct. 2020, doi: 10.3389/fphys.2020.569050.

[25] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism," *Appl. Sci.*, vol. 10, no. 17, p. 5841, Aug. 2020, doi: 10.3390/app10175841.

[26] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.

[27] W. Mumtaz and A. Qayyum, "A deep learning framework for automatic diagnosis of unipolar depression," *Int. J. Med. Inf.*, vol. 132, p. 103983, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103983.

[28] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal Measurement of Depression Using Deep Learning Models," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, Mountain View California USA: ACM, Oct. 2017, pp. 53–59. doi: 10.1145/3133944.3133948.