*International Journal of Computational and Experimental Science and ENgineering (IJCESEN)*

*Vol. 11-No.2 (2025) pp. 2445-2458*
*http://www.ijcesen.com*

***ISSN: 2149-9144***



**Research Article**

# A Hybrid Deep Learning Framework and Dwarf Mangoose Optimized Layers for an Effective Depression Classification

**K. Neeraja[1*], G. Narsimha[2]**

[1]Research Scholar, Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana, India
* **Corresponding Author Email:** neeru5809@gmail.com - **ORCID:** 0009-0000-2659-1679

[2]Professor and Principal, Computer Science and Engineering, JNTUH University College of Engineering Sultanpur, Telangana, India
**Email:** narsimha06@jntuh.ac.in - **ORCID:** 0000-0002-0143-8122

**Article Info:**

**DOI:** 10.22399/ijcesen.1738
**Received :** 02 February 2025
**Accepted :** 05 April 2025

**Keywords :**

Depression,
Adapted Dilated Convolutional Network,
Dwarf Mongoose Model,
Hybrid Ensemble Model.

**Abstract:**

Depression is considered to be one of the dangerous diseases that affects physical state of human and even causes the fatal end to the patients. The depression leads to the anxiety disorders, bipolar disorders and at the same time may hit the person's mind set to the suicide thoughts. Hence it is considerably demanding task to recognise the individuals with mental conditions. Traditionally, depression detection was done through patient's interview and PHQ scores, but these traditional methods produce the accuracy which has very little effect on the diagnosis and treatment process. With the advent of machine (ML) and deep learning (DL) models, depression detection has reached its new dimensional path but still diagnosis performance and computational overhead remains to be real bottleneck for achieving its own strength of classification and early diagnosis. To solve this aforementioned problem, this research paper proposed hybrid ensemble of deep learning models and optimized training networks. Proposed framework consists of three components: first, adaptive dilated convolutional networks in which the model is trained with text and audio features, third is Bi-GRU Networks a finally the learning layers are optimized by DwarfMangoose Model to attain the better classification of depressions. The recommended model is examined and evaluated by utilizingDAIC-WOZ database and performance metrics such as accuracy, precision, recall, specificity and F1-score are measured and examined with the varied state-of-art learning procedures. The results demonstrate the recommended model has provided the optimal solution in detecting the depressions and produced the accuracy of 0.98, precision of 0.972, recall of 0.98, specificity of 0.98 and F1-score of 0.987 respectively. Experimental findings have proved that the proposed model has produced the promising results that improvises the clinical treatment and overcomes the fatal fears of the patients caused by the depression.

## 1. Introduction

Depression is a prevalent health condition and ranks among the most widespread mental illnesses, impacting millions of individuals worldwide. It is recognized as a severe ailment that not only impairs a person's psychological well-being but can also inflict physical consequences on the individual. The intensity of depression is often assessed based on the patient's mental health status [1]. Commonly observed mental health issues include anxiety disorders, restlessness, sleep disturbances, eating irregularities, addiction-related conditions, depression, trauma, and stress-induced disorders [2]. Depression is a mental health disorder where individuals experience persistent feelings of hopelessness, lack of motivation, mood instability, and diminished interest in routine physical, mental, and social activities. These symptoms often result in emotional distress and physical changes within the body. This condition notably impairs a person's ability to learn, triggers mood variations, and frequently reduces their overall work productivity. The symptoms and intensity of depression vary depending on its severity [3]. In severe cases, brain activity slows down, leading to increased production of cortisol, a hormone that inhibits neuron growth in the brain. This significantly disrupts cognitive

processes and, in extreme situations, may result in suicidal tendencies. Depression manifests in various forms, including clinical depression, bipolar disorder, dysthymia, seasonal affective disorder, among others [4].

### 1.1 Background

Numerous Studies have been utilized on the early detection of depression detections. Conventional approaches for diagnosing depression often rely on interview-based systems and structured evaluation tools like the Hamilton Depression Rating Scale. Self-assessment instruments, such as the Beck Depression Inventory and the PHQ-8 questionnaire, are also widely utilized. Additionally, structured clinical interviews are employed to assess symptom severity and identify common behavioural patterns in individuals experiencing depression [5–7]. However, traditional treatments, including psychotherapy and pharmacological interventions, are frequently criticized for being time-intensive, costly, and less effective in delivering long-term solutions. As the depression affected patients are increased, artificial intelligence (AI) techniques such as machine and DL algorithms are witnessing the increase in designing the automatic expert systems for an early detection of depression [8-11].

### 1.2 Problem Statement

As mentioned above, DL algorithms such as Convolutional Neural networks (CNN) [12], Long Short-term memory (LSTM) [13], Gated Recurrent neural networks (RNN) [14] are mostly widely used for the design of early detection of depressions. A significant limitation of current methods for detecting depression lies in their heavy reliance on extensive patient information. These techniques often require comprehensive data, including the patient's background, medical history, and details of any past traumas, to identify symptoms of depression effectively. Furthermore, many of these approaches demand continuous monitoring of patient activities to determine whether an individual is experiencing depression. Hence the deep learning model needs more improvisation to handle the larger datasets without creating the over fitting problem.

### 1.3 Proposed Solution

Motivated by the aforementioned problem, this paper recommends the hybrid DL model by the powerful integration of Dilated Convolutional Layers, Bi-Gated Recurrent Neural networks and Dwarf Mangoose Optimized Dense Classification layers for an early detection of depression taking the multiple data features. The proposed hybrid model takes the different features and classifies the depressed patients accurately when compared with the traditional models.

### 1.4 Contribution of the Paper

The significant contribution of the research paper is pursues:

1. The paper recommends the hybrid ensemble of Dilated Convolutional Layers (DCL) and Bi-GRU networks for an effective extraction of the features from the different set of datasets.
2. The paper introduces the novel dwarf mongooses (DM) optimization model to tune the hyper-parameters of the classification networks to attain the optimal detection of depressions.
3. The extensive experimentation is utilized by the DAIC-WOZ datasets which consists of different data features such as text, audio and video input features. The experimental outcomes depict the superiority of the recommended approach.

### 1.5 Organization of the Paper

The structure of the paper is organized as follows as: 1) **Section- 2** presents the related works by the more than one authors. 2) The dataset descriptions, proposed deep learning frameworks with its principle of working are illustrated in **Section-3.** The experimental evaluation, performance analysis, results discussions are presented in **Section-4**. Finally, the paper is concluded with the future enhancement in **Section-5.**

## 2. Related Works

Beniwal and Saraswat (2024) [15] developed a hybrid BERT-CNN approach for depression detection on social media utilizing multimodal data. They created a dataset containing text, emoticons, and image data, and proposed a model consisting of three parts: a textual BERT model for text and emoticon data, a CNN for image data, and a hybrid BERT-CNN model for both text and images. The BERT model achieved 97% accuracy for text data, while CNN attained 89% accuracy for image data. While the model showed promising results, it is limited by the complexity of integrating multiple data types. Figure 1 shows dataset distribution used in the proposed model training.

Yenugutalaa (2024) [16] explored variedML and DLprocedures for depression detection, including CNN, RNN, Support Vector Machine (SVM), Random Forest, Logistic Regression, and Naive Bayes. Using a dataset of tweets from Kaggle, the RNN model attained the best accuracy The study

demonstrates high performance metrics, but its reliance on a single data source and the potential for overfitting based on specific dataset characteristics pose challenges. Tejaswini et al. (2024) [17] introduces a hybrid DL model called "Fasttext Convolution Neural Network with Long Short-Term Memory (FCL)" for depression detection from text. The model utilizes fasttext embedding for better text representation, CNN for extracting global information, and LSTM for capturing local features with dependencies. The FCL model outperformed state-of-the-art methods in detecting depression with high accuracy on real-world datasets. However, they do not address potential limitations such as computational requirements and generalizability across different types of text data. Vandana et al. (2023) [18] developed a hybrid model for depression detection utilizing DL approaches, combining textual and audio features from patient responses. Using the DAIC-WoZ database, they implemented textual CNN, audio CNN, and hybrid LSTM and Bi-LSTM models. The audio CNN model achieved 98% accuracy, outperforming the textual CNN model (92% accuracy). The Bi-LSTM model demonstrated a better learning rate with 88% accuracy and 78% validation accuracy. Despite the paper's encouraging findings for multimodal depression detection, limitations may occur due to the distinct characteristics of the DAIC-WoZ database. Khafaga et al. (2023) [19] presented a DL Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN) for classifying depression data from Twitter. The proposed MDH-PWO architecture achieved 99.86% accuracy, outperforming existing methods such as CNN, SVM, and MDL. The study showcases high accuracy, along with overfitting issues and challenges in adapting to other social media platforms or text sources. Amanat et al. (2022) [20] proposed a productive model implementing LSTM with two hidden layers and large bias, combined with a RNN with two dense layers, to predict depression from text. The framework achieved 99.0% accuracy, outperforming frequency-based DL models while reducing the false positive rate.

Kalpana et al. (2025) [21] suggested an explainable AI-based gait analysis model with wearable IoT for human activity recognition. Their method increases interpretability in smart healthcare, but its applicability to heterogeneous patient conditions and real-time deployment issues is not yet explored.

Nadeem et al. (2022) [22] introduced a novel Sequence, Semantic, Context Learning (SSCL) classification framework with a self-attention mechanism for depression detection from textual data. The framework utilizes GloVeembeddings, LSTM, CNN, GRUs, and self-attention to capture various aspects of tweets. The SSCL framework achieved 97.4% accuracy for binary labeled data and 82.9% for ternary labeled data. It also demonstrated strong performance on unseen data and cross-domain validation. Despite its impressive results, the research did not extensively address potential limitations such as computational complexity and scalability to larger datasets. Kalpana et al. (2024) [23] proposed a deep reinforcement learning-based task offloading framework for edge-cloud computing. Their method optimizes resource allocation, but its scalability and real-time efficiency in highly dynamic cloud environments need to be validated.

## 3. System Overview

Figure 2 presents the system architecture for the recommended approach. As shown in Figure 2, the recommended framework consists of four different phases such as Data Collection Process, Data pre-processing, Hybrid Deep learning model and finally early classification of depression. The proposed hybrid architecture consists of two components such as adaptive dilated convolutional layers and Bi GRU networks followed by the optimized learning networks by dwarf mongoose algorithm. In the first component, audio files are converted into MFCC co-efficients and feed to the Bi-Gated Recurrent Units (Bi-GRU networks). In the second part, textual features are used to train the adaptive dilated convolutional networks. The features from the two components are then fused by the feature fusion model. Finally, the fused features are then act as the input to the optimized learning networks to detect the depression of the patients. In the suggested framework, audio and text features are processed using a combination of CNN and LSTM architectures, which work together to produce binary outputs: Depressed or Not-Depressed. This model is specifically designed for identifying depression and can autonomously determine whether an individual exhibits signs of depression or not. The detailed description of every component are presented as follows

### 3.1 Materials and Methods

The DAIC-WOZ (Distress Analysis Interview Corpus-Wizard of Oz) database is a comprehensive resource designed to aid the development of automated depression detection systems. This dataset contains an extensive collection of data, including responses from individuals experiencing recordings, and textual responses to questionnaires. Its utilization is driven by the abundance of data available across these modalities.

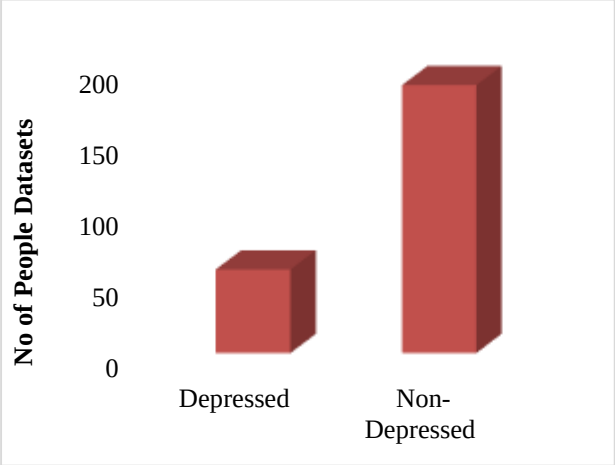


***Figure 1.*** *Dataset Distribution used in the Proposed Model Training*

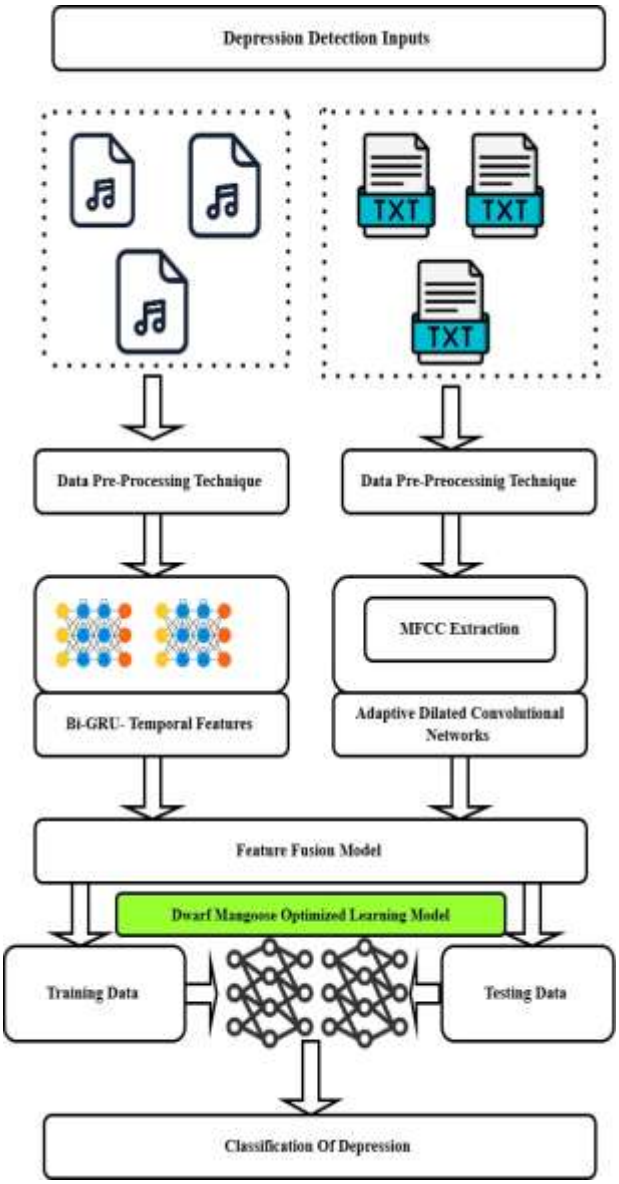


***Figure 2.*** *Proposed Architecture for the Overall Framework*

depression. The database includes multimodal information in the form of audio recordings, video The DAIC-WOZ dataset can be accessed through the Kaggle repository [24]. In this research, audio files and text files are taken as the inputs for training and validating the recommended approach.

### 3.2 Proposed MFCC Feature Extraction for Audio Files

In this study ,advanced methodologies implied in speech and voice assessment which can be suitable for an effective audio analysis. The recommended study employs Mel-Frequency Cepstral Coefficients (MFCCs) and harmonic ratios as key distinguishing features, emphasizing frequency-based attributes to enhance the classification process. MFCCs are computed by deriving the power spectrum using the Fast Fourier Transform (FFT) and subsequently converting it to the mel scale, which represents a perceptual pitch scale reflecting the human auditory system's non-linear response to various frequencies. The power spectrum data is then passed through triangular filters uniformly distributed along the mel scale. The logarithmic transformation applied in this process simulates the human ear's sensitivity to variations in loudness across different frequency ranges.The discrete cosine transform (DCT) is then applied to the logarithmic power spectrum, aiming to reduce correlations among values and highlight the most significant features. This transformation generates coefficients that represent the cepstral domain, encapsulating the signal's key attributes. Adopting 13 Mel-frequency cepstral coefficients (MFCCs) conforms to recognized standards in audio analysis, effectively balancing the need for a detailed depiction of the spectral envelope with the goal of minimizing redundancy (figure 3-4). This approach addresses the uniformity of sound representation. As a result, MFCCs provide a detailed and nuanced snapshot of vocal performance during a given task, leveraging a nonlinear frequency scale to offer a comprehensive description of perceptual aspects of human speech. In the cepstral domain, details regarding the rate of change across various frequency bands are captured due to the calculation of cepstral coefficients through the Fourier transform of the log-transformed spectrum. This distinctive trait of cepstral features offers significant benefits, as it enables the differentiation between the contributions of the source (vocal cords) and the filter (vocal tract) in speech signals. To extract the Mel Co-efficients using the above methods has been deployed using the MFCC libraries in the python programming languages**.** Table 1 to 3 presents the MFCC co-efficient that are utilized for training the recommended model.

***Table 1.*** *MFCC Coefficients extracted for the different voice samples of hypokinetic dyshonia*

| MFCC-1 | MFCC-2 | MFCC-3 | MFCC-4 | MFCC-5 | MFCC-6 | MFCC-7 | MFCC-8 | MFCC-9 | MFCC-10 | MFCC-11 | MFCC-12 | MFCC-13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -136.557 | 87.9028 | 46.2129 | 3.13008 | 29.5048 | 1.42611 | 20.6897 | 0.65477 | 15.4370 | 0.09575 | 12.2242 | -0.10136 | 10.1931 |
| -204.009 | 88.1811 | 47.6436 | 3.4318 | 30.9214 | 1.7334 | 22.1028 | 0.96290 | 16.8476 | 0.40409 | 13.6318 | 0.20718 | 11.5973 |
| -265.945 | 30.0022 | 41.6593 | 15.6005 | 15.9383 | 13.5374 | 14.2126 | 12.8575 | 22.3491 | 3.10506 | 11.8919 | 6.20416 | 0.86194 |
| -0.91224 | 10.2501 | 22.4634 | 12.5521 | 10.6689 | 13.5754 | 5.21625 | 11.5629 | 15.6452 | 7.81451 | 9.70887 | 5.76354 | -0.27693 |
| 71.88125 | 2.38821 | 11.0641 | 5.63598 | 4.06610 | 5.43475 | 0.65175 | 2.45894 | 9.36693 | 10.7018 | 14.4874 | 12.7882 | 4.56596 |
| 80.23127 | -7.40041 | -3.00866 | -2.71379 | -2.61805 | -7.79448 | -2.39965 | -4.8528 | 2.57649 | 8.18816 | 11.7001 | 8.62647 | 0.36212 |
| 82.7937 | -4.9182 | -2.8281 | -3.17682 | -3.02311 | -9.49494 | -8.72688 | -8.68509 | -0.02228 | 4.60878 | 9.35640 | 7.39344 | -1.69381 |
| 90.98094 | 2.380652 | 0.913503 | -2.33838 | 4.181436 | 3.319114 | 0.160641 | -3.986 | 6.934484 | 4.019251 | 0.065877 | 1.520992 | -1.21386 |
| 89.63231 | 2.395089 | 0.998017 | 2.511336 | 9.92436 | 3.177996 | 2.889478 | -1.73769 | 3.656701 | 1.971443 | 0.15245 | -2.48936 | -2.49449 |
| 91.60279 | 1.194155 | -0.76251 | 1.138315 | 9.38246 | 3.241541 | 3.736206 | -2.58256 | 0.77301 | 1.66985 | 4.037854 | 1.09147 | -0.25271 |
| 85.20509 | -5.40553 | -3.38717 | 1.240818 | 7.791197 | -0.56688 | -0.33787 | -6.12673 | -3.18445 | -2.73867 | 0.75711 | 0.257689 | -1.93569 |
| 80.58945 | -5.63263 | -3.8292 | -0.16862 | 6.58735 | -2.01299 | 0.153615 | -3.96277 | 2.135773 | -0.24012 | 3.067284 | 5.113432 | -1.81989 |
| 82.36979 | -4.05794 | -7.32285 | -6.53906 | 4.220176 | -1.74477 | -1.20826 | -0.99251 | 5.385958 | 0.125405 | 2.889936 | -0.72966 | -8.52589 |
| 86.97674 | -3.40706 | -5.49574 | -6.29646 | 1.774815 | 0.403348 | -1.90903 | -0.16928 | 0.105797 | -1.8794 | 1.053323 | -9.32057 | -9.36436 |

***Table 2.*** *MFCC Coefficients extracted for the different voice samples of reflux laryngitis*

| MFCC-1 | MFCC-2 | MFCC-3 | MFCC-4 | MFCC-5 | MFCC-6 | MFCC-7 | MFCC-8 | MFCC-9 | MFCC-10 | MFCC-11 | MFCC-12 | MFCC-13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 91.48146 | 0.68366 | 2.698182 | -1.40848 | 2.680625 | 0.190965 | -1.93138 | -1.67583 | -1.42352 | -3.26002 | 6.114356 | 0.929124 | 2.202178 |
| 93.93761 | 0.4378 | -2.1392 | -5.96047 | 3.392513 | 5.003177 | 0.441009 | -2.63749 | -0.59316 | -0.75739 | 3.246762 | -0.10011 | 4.56648 |
| 93.89419 | 0.50875 | -4.24729 | -10.1068 | -1.98071 | 1.327945 | 4.671668 | -3.54924 | -3.05958 | -2.16874 | 2.224508 | 0.303717 | 5.215704 |
| 95.6907 | 3.789573 | -0.00964 | -7.83308 | -0.99889 | 2.022135 | 6.412212 | -5.51323 | -3.98255 | -4.21977 | -0.08907 | 0.164835 | 4.704379 |
| 95.09955 | 6.323491 | 2.699694 | -7.53554 | 1.655296 | 2.547106 | 5.159884 | -1.3164 | -0.51482 | -5.56153 | 1.806301 | -0.8205 | 8.147171 |
| 91.84375 | 5.975073 | 3.561192 | -4.66012 | 3.335255 | 6.507042 | 7.924284 | 1.92456 | -0.00509 | -5.67588 | 2.663113 | 2.884445 | 5.009964 |
| 90.09317 | 2.84396 | 5.109411 | -0.14552 | 6.536883 | 7.605429 | 9.478476 | -1.23974 | -6.00955 | -6.8483 | 4.746659 | 4.579631 | 2.01576 |
| 88.4081 | -3.30225 | -3.48283 | -2.63882 | 7.679319 | 7.889988 | 8.594012 | -3.48355 | -7.27104 | -4.27738 | 2.97416 | -0.87667 | 2.385092 |
| 91.98953 | -4.18677 | -8.01927 | -6.79491 | 3.748691 | 7.797329 | 9.314379 | -1.59296 | -8.45745 | -5.8425 | 1.662824 | 2.503189 | 6.980917 |
| 93.6742 | 1.091373 | -0.97927 | -6.27153 | 0.086835 | 4.955744 | 0.785921 | -6.37688 | -7.38315 | -3.58509 | 7.972229 | 6.932082 | 2.894475 |
| 93.12556 | 4.447236 | 3.662921 | -5.08253 | -1.00526 | -0.08067 | -0.37082 | -5.81599 | -1.07317 | -0.1219 | 4.497883 | 2.074391 | 3.936482 |
| 89.49228 | 3.027795 | 1.399508 | -6.62571 | 2.535962 | 4.384495 | 4.057686 | -4.06641 | -4.08099 | -6.18707 | -2.08505 | -5.91245 | 2.154335 |
| 89.50396 | 1.218399 | 2.436895 | -3.35732 | 5.272279 | 8.18457 | 5.604684 | -3.12141 | -1.57856 | -5.76492 | -4.87502 | -6.43603 | -0.70068 |
| 94.11093 | 7.046132 | 2.183645 | -3.5092 | 2.786134 | -2.00193 | 4.574269 | -7.33736 | -5.21181 | -4.05376 | 1.832304 | -0.04219 | 1.090493 |

***Table 3.*** *MFCC Coefficients extracted for the different voice samples of Healthy voices*

| MFCC-1 | MFCC-2 | MFCC-3 | MFCC-4 | MFCC-5 | MFCC-6 | MFCC-7 | MFCC-8 | MFCC-9 | MFCC-10 | MFCC-11 | MFCC-12 | MFCC-13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 95.54342 | -0.60419 | 0.001952 | 1.766276 | 7.172626 | -3.82589 | -1.4827 | -5.56376 | 1.615592 | 6.562648 | 6.836293 | -5.73365 | -3.51436 |
| 91.86496 | -3.76409 | 0.112926 | 4.123205 | 8.638082 | -5.48514 | -1.26807 | -0.53159 | 5.281814 | 3.804416 | 7.609248 | 0.725268 | -0.83909 |
| 91.68339 | 1.872721 | -2.88371 | -3.85844 | 5.143834 | -7.36987 | -5.36551 | -6.47391 | 1.079441 | 0.469132 | 1.718446 | -6.09787 | 1.903828 |
| 93.98507 | 4.922437 | -0.46217 | -4.10127 | 4.726997 | -6.85205 | -4.55062 | -11.0236 | -1.07257 | -0.65855 | -3.92207 | -11.091 | -1.07993 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 90.92912 | 2.607314 | -0.06094 | -0.7445 | 11.59087 | 2.041338 | 0.365609 | -7.80323 | -0.79172 | -4.29813 | -5.515 | -8.69073 | -3.72506 |
| 92.06683 | 2.361665 | 3.144542 | 1.450806 | 11.21821 | 3.128822 | 1.70862 | -8.4715 | -3.84775 | -3.03254 | 0.423664 | -3.44118 | 0.564664 |
| 93.75628 | -1.6305 | 1.917774 | 3.151506 | 8.260822 | -0.7512 | -2.27236 | -6.47678 | -3.54682 | -3.71993 | 1.805558 | -1.58678 | -0.46992 |
| 93.1894 | 1.709203 | 0.821961 | 2.516522 | 9.755192 | -2.39587 | 0.35517 | -4.0503 | 1.188923 | -2.62481 | 3.546936 | 0.255482 | -2.4023 |
| 93.27333 | 2.726733 | -2.65746 | -2.91864 | 8.743229 | -4.24943 | -3.70714 | -7.45994 | 3.984869 | 0.246267 | 2.280527 | 0.247707 | 1.267272 |
| 93.34255 | 6.635552 | -3.26715 | -8.62687 | 6.149344 | -1.9074 | -5.7267 | -6.15541 | 4.643708 | -0.12608 | 0.588463 | -3.17762 | 3.143801 |

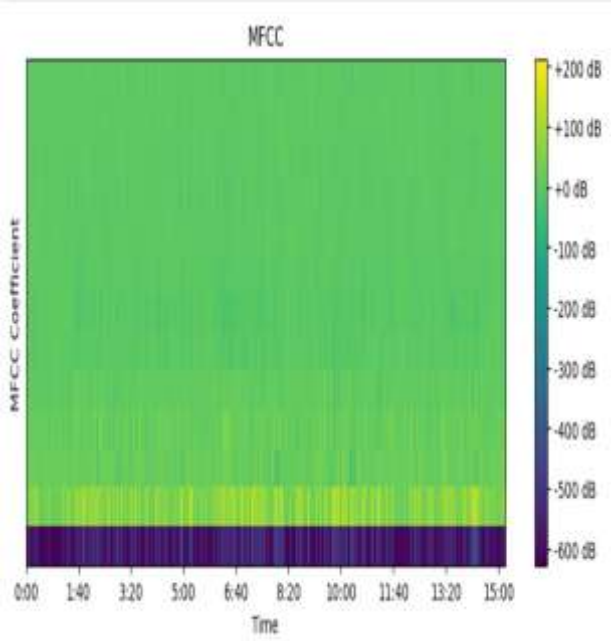


***Figure 3.*** *Sample MFCC Representation for the Depressed Audio Files*

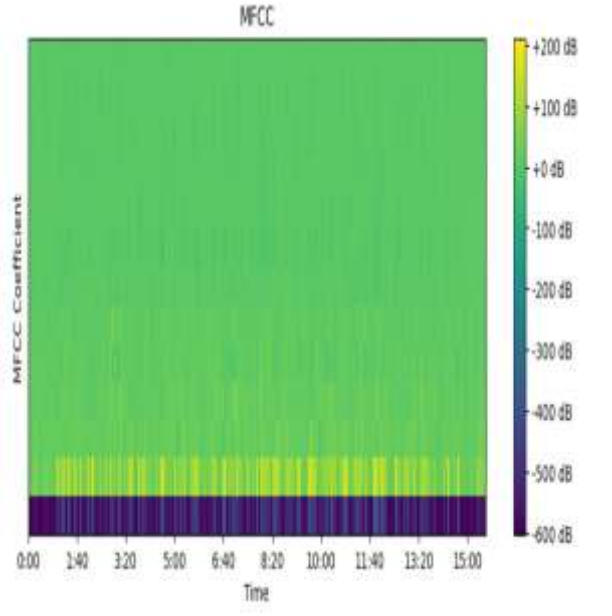


***Figure 4.*** *Sample MFCC Representation for theNon-Depressed Audio Files*

## 3.3 Adaptive Dilated Convolution Networks

The recommended network conducts multi-scale feature extraction on the input feature maps using dilated convolution operations with adjustable dilation factors, followed by concatenating the final multi-scale features. In contrast to other architectures, the fusion of multi-scale features is performed without relying on up-sampling, which typically leads to higher computational costs. As shown in Figure 2, recommended model is constructed using dilated convolutional networks including four DCR in which convolutional kernel size is 3x3 with adaptive dilation rate is set to 1,2,4,8 and the channel is set to 64. The receptive field in each layer is calculated using mathematical expression

$$J_k = J_{k-1} + (|\, G_{k-1}) * \sum_{t=1}^{K} Tt \qquad (1)$$

Where $J_k$ is the receptive fields K is Kernel size and t is the stride length ad pooling layers.Furthermore, features from the each attention layers are calculated as

$$f1 = Softmax\,(F(1)) \qquad (2)$$

Hence overall features are given by concatenating all the features from dilated convolutional maps are given by

$$F = Concat(f1, f2, f3 \ldots\ldots) \qquad (3)$$

The concatenation operation independently determines the weight for each feature, where *fi* represents the feature map obtained from each layer, and *F* refers to the combined feature passed to the subsequent NRB.

## 3.4 Bi-GRU networks For Textual Feature Extraction

The Gated Recurrent Unit (GRU) is regarded as a highly intriguing variant of the LSTM network. The concept was described by Chung et al. [25], aiming

to merge the forget gate and input vector into a unified vector. This architecture effectively handles long sequences and preserves long-term dependencies. Moreover, the GRU offers significantly reduced complexity compared to LSTM networks.

**Upcoming equations are described by Chung to depict the attributes of GRU**

$$h_t = (1 - x_t) \odot h_{t-1} + x_t \odot h_t \quad (4)$$

$$\widetilde{h_t} = g(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h \quad (5)$$

$$z_t = \sigma(W_h x_t + U_z h_{t-1} + b_z) \quad (6)$$

$$r_t = \sigma(W_h x_t + U_r h_{t-1} + b_r) \quad (7)$$

The entire GRU equation is depicted by

$$P = GRU\left(\sum_{t=1}^{n}[x_{t,} h_{t,} z_{t,} r_t\big(W(t), B(t), \eta(tannh)\big)\right) \quad (8)$$

At the present time step, $x_t$ depicts the input feature, while $y_t$ indicates the corresponding output state. The output of the module at this time step is denoted as $h_t$. Additionally, $z_t$ and $r_t$ represent the update and reset gates. The weight matrix at this specific instant is labelled as W(t), and the bias weights at this time step are represented by B(t).

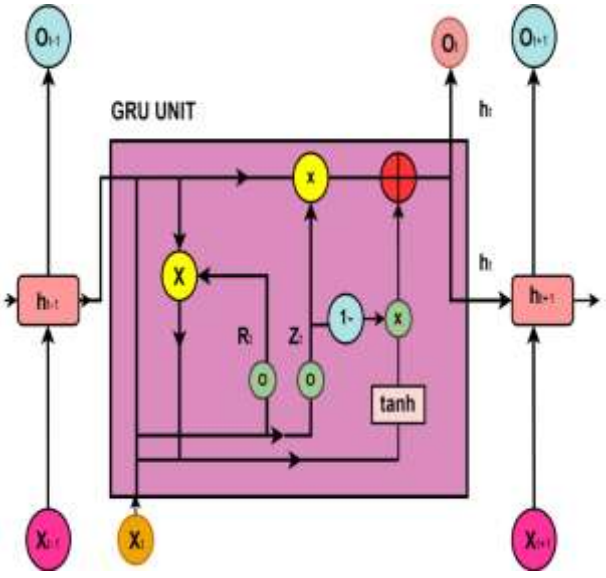


***Figure 5.*** *GRU Networks used for the Feature Extraction using the textual data features*

Figure 5 is GRU networks used for the feature extraction using the textual data features. To extract relevant information from various sources within the dataset, BiGRU (Bidirectional Gated Recurrent Unit) networks are utilized, consisting of both forward and backward GRU cells. The mathematical formulation of the BiGRU network is represented in Equation (11). This BiGRU network effectively captures temporal features, which encompass diverse information beneficial for classification. However, the extracted features can introduce excessive diversity, potentially increasing training time and leading to overhead in the classification layer. To mitigate this computational burden during classification, Self-attention layers are incorporated among the BiGRU network and the classification layer. The input features from the BiGRU network are passed through a softmax layer to generate attention-based features, as indicated in Equation (8), which are subsequently fed into the feed-forward classification layers.

$$\boldsymbol{P(F)} = GRU\left(\sum_{t=1}^{n}[x_{t,} h_{t,} z_{t,} r_t\big(W(t), B(t), \eta(tannh)\big)\right] \quad (9)$$

$$P(B) = GRU\left(\sum_{t=1}^{n}[x_{t,} h_{t,} z_{t,} r_t\big(W(t), B(t), \eta(tannh)\big)\right] \quad (10)$$

Combining the Equation (9) and (10)

$$P(BiGRU) = P(F) + P(B) \quad (11)$$

## 3.5 Feature Fusion Layers

To fuse the features from the different layers of networks, flatten layers are employed for collecting and combining the features.

## 3.6 Dwarf Mongoose Based Deep Learning Networks

The dense classification network whose hyper-parameters are tuned by the proposed dwarf mongoose algorithm to achieve the higher performance and less computational complexity. The dwarf mongoose (*Helogale*), native to Africa's semideserts and savanna bushlands, is often found in environments rich in termite mounds, rocks, and hollow trees that serve as their shelters. Recognized as Africa's smallest carnivorous mammal, it measures around 47 cm in total body length and

weighs approximately 400 g in adulthood. These mongooses form tightly-knit family groups governed by a matriarchal structure, with an alpha male and female pair leading the group and maintaining a lifelong bond. In this hierarchy, females dominate males, and within each age group, younger individuals hold precedence over their elder siblings. The cooperative behavior and selflessness observed in these family units are unmatched among mammals. Responsibilities within the group, such as serving as sentinels, caring for offspring, and defending against predators or rival intruders, are allocated based on age and sex. Comprehensive studies have documented the group's roles, hierarchical structures, and interpersonal dynamics. Dwarf mongooses are highly territorial animals, utilizing their cheek and anal glands to leave scent marks on vertical or horizontal surfaces within their domain. These scent markings provide a sense of familiarity, fostering confidence and reducing anxiety among the group members. All individuals in the family unit participate in marking activities, with the frequency of contributions linked to their hierarchical rank. Research conducted on dwarf mongooses in the Taru Desert revealed that their territories are characterized by features such as termite mounds, rocky formations, and hollow trees. On average, a family required 21.8 days to traverse and re-mark their entire territory. The study further indicated that group size influences the pace of territory coverage, foraging behavior, and the use of sleeping mounds, although the boundaries of their territory remain constant. Dwarf mongooses strategically maintain the smallest defendable area that provides adequate resources for the group. The dwarf mongoose exhibits unique behaviors and adaptations tied to its territorial nature and strategies for avoiding predators. Lacking a powerful bite, these mongooses employ a skull-crushing technique to kill prey, using the prey's eye as a reference point. This specific method limits the size of their prey and precludes cooperative hunting of larger targets. Consequently, this constrained method of capturing prey has a notable impact on their social dynamics and ecological strategies, requiring adjustments to ensure adequate nourishment for their group. The mongoose compensates through two primary behavioral strategies aimed at addressing these nutritional challenges. The dwarf mongoose exhibits a restricted prey-catching behavior, targeting only small prey, which are insufficient for sharing among adults but can be used to feed the young. Their diet primarily includes arthropods, although they occasionally capture small mammals, geckos, and small birds. The prey availability for dwarf mongooses is scattered and unpredictable, necessitating extensive searching to meet their dietary needs. To adapt to this challenge, dwarf mongooses have adopted a seminomadic lifestyle, traveling significant distances during foraging and seldom revisiting previously utilized sleeping mounds. This strategy prevents overexploitation of specific areas, minimizes prey scarcity, and ensures balanced hunting across their territory.

The group forages collectively, maintaining cohesion through vocal signals, typically a soft nasal "peep" sound at a frequency of 2 kHz, produced by the alpha female. The daily distance travelled by the group is influenced by factors such as group size, the presence of offspring, and interruptions caused by predator evasion. The alpha female plays a central role in initiating foraging activities, setting the direction and range of movement, and selecting sleeping mounds for the group.

The absence of prey provision to lactating females or younglings in dwarf mongooses has significantly influenced their social structure, particularly in parenting practices, resulting in the following adaptive behaviour:

## 3.7 Allow parenting or Babysitting

A portion of the population, comprising both males and females, acts as babysitters for the young. These individuals stay with the offspring until the group reconvenes at midday, at which point they swap roles with other group members who then resume foraging. Dwarf mongooses do not construct nests for their young but instead relocate them between different sleeping mounds. When the young join the group, they are underdeveloped, with sparse fur and limited stamina, making them unable to travel long distances. This restricts the group's daily foraging range and limits overall movement.

In essence, dwarf mongooses are unable to hunt or capture substantial prey that could sustain the entire group. The lack of a lethal bite and organized group hunting has led them to adopt a social system where individuals primarily fend for themselves. This unique system involves constant relocation. Their nomadic behavior prevents overuse of specific areas and ensures thorough exploration of their habitat, as they avoid returning to previously used sleeping mounds.

## Population initialization

The DMO optimization process starts by randomly initializing the mongoose candidate population XX, as described in Equation (12). This population is created within the defined range of the problem, bounded by the upper limit (UB) and lower limit (LB).

$$X = \begin{bmatrix} x_{1,1}x_{1,2}\cdots x_{1,d-1}x_{1,d} \\ x_{2,1}x_{2,2}\cdots x_{2,d-1}x_{2,d} \\ \vdots\ x_{i,j}\ \vdots \\ x_{n,1}x_{n,2}\cdots x_{n,d-1}x_{n,d} \end{bmatrix} \quad (12)$$

The current set of candidate populations is created through a random generation process as per Equation (12). Here, $x_{ij}$ represents the location of the jth dimension for the ith population member, while PP stands for the total number of populations. Additionally, dd specifies the dimensionality of the given problem.

$$x_{i,j} = unifrnd(VarMin, VarMax, VarSize) \quad (13)$$

A uniformly distributed random variable is denoted as r, while LB and UB signify the lower and upper bounds of the problem, respectively. Additionally, D represents the number of decision variables or the dimensionality of the problem. The optimal solution identified at any iteration is regarded as the best solution achieved up to that point.

## The DMO model

The recommended DMO approach mimics the compensatory behavioural strategies observed in dwarf mongoose populations. These strategies involve adaptations such as limiting the size of prey, establishing social roles (like babysitters), adopting a seminomadic lifestyle, and more. In implementing this algorithm, the dwarf mongoose's social structure is categorized into three main groups: the alpha group, scouts, and babysitters. Each subgroup plays a vital role in enabling compensatory behaviours that support their seminomadic lifestyle within a territory that can sustain the entire community.

Dwarf mongooses are known for their coordinated foraging and scouting activities, which are conducted collectively. In this model, the same group undertakes both foraging and scouting for new mounds concurrently. The alpha group begins by searching for food while simultaneously identifying new sleeping mounds, which are used once the criteria for babysitter rotation are fulfilled. The simulation evaluates the average value of sleeping mounds in each iteration, guiding the population's next movement based on the computed value, as detailed in Equation (14). This nomadic movement minimizes the overuse of specific areas and promotes exploration across the entire territory, ensuring no previously used mound is revisited. The optimization steps within the DMO algorithm are divided into three distinct phases, which are illustrated as

### Alpha Group

Once the initial population is generated, the fitness score of each individual solution is evaluated. The likelihood of selecting each solution, determined by its fitness score, is calculated using Eq. (14). Based on this probability, the alpha female is identified as the most fit candidate.

$$\alpha = \frac{fit_i}{\sum_{i=1}^{n} fit_i} \quad (14)$$

$$\boldsymbol{X_{i+1}} = X_i + phi * peep \quad (15)$$

The variable φ is a random value drawn from a uniform distribution over the range [-1, 1]. Following each iteration, the configuration of the sleeping mound is persistent by the expression in Equation (16).

$$sm_i = \frac{fit_{i+1} - fit_i}{max\{|fit_{i+1}, fit_i|\}} \quad (16)$$

### Scout Group

The scouts search for the next sleeping mound, as mongooses are known to avoid returning to their previous resting sites, ensuring continuous exploration. This scouting process occurs concurrently with foraging. The movement is evaluated as a success or failure in discovering a new sleeping mound, meaning the mongooses' overall performance influences this activity. Essentially, if the group forages sufficiently far, they are likely to find a new resting location. Equation (17) represents the behavior of a scout mongoose.

$$\boldsymbol{X_{i+1}} = \begin{cases} X_i - CF * phi * rand * [X_i - \vec{\boldsymbol{M}}] & if\ \varphi_{i+1} > \varphi_i \\ X_i + CF * phi * rand * [X_i - \vec{\boldsymbol{M}}] & else \end{cases} \quad (17)$$

The variable *rand* represents a random value between 0 and 1. The formula $CF = \left(1 - \frac{iter}{Max_{iter}}\right)^{\left(2\frac{iter}{Max_{iter}}\right)}$ is a parameter that influences the collective and volitional motion of the mongoose group, with its value decreasing gradually as the iterations progress. The equation $\vec{\boldsymbol{M}} = \sum_{i=1}^{n} \frac{X_i \times sm_i}{X_i}$ calculates the vector directing the mongoose's movement toward a new resting location. This formulation helps to model the directional change in the mongoose's position based on the collective behavior of the group. The babysitters typically form the subordinate members of the group, staying with the young ones while the

alpha female (mother) leads the rest of the group in foraging activities. To ensure balance, the babysitters are rotated frequently. The alpha female typically returns during midday and in the evening to nurse the young. The number of babysitters in the group is proportional to the total population size. These dynamic influences the approach by lowering the population size in proportion to the percentage allocated for babysitters. To simulate this, we reduce the total population by a percentage corresponding to the babysitter count. Additionally, the babysitter exchange parameter is used to reset the information related to scouting and food sources, which was previously held by the group members being replaced. Babysitters are assigned a fitness weight of zero, which lowers the average weight of the alpha group during subsequent iterations. This leads to a reduction in group movement, which in turn promotes the process of exploitation.

### 3.8 DMO Tuned Classification Networks

The dense network weights utilized by the classification layers are fine-tuned using the straightforward DWO approach. Initially, the training network adopts randomly chosen hyperparameters. The effectiveness of the suggested method is determined using the fitness value outlined in Equation (18). For every iteration, the hyperparameters are calculated using Algorithm-1. The iteration terminates as soon as the fitness function aligns with the condition specified in Equation (18).

$$Fitness\ Function = \{Max(Accuracy, Precision, Recall, F1-score) \quad \textbf{(18)}$$

The classification layer proposed in this work operates efficiently and with minimal computational requirements to identify normal and heart disease conditions. Algorithm 1 outlines the operational mechanism of the suggested classification layer. The proposed framework incorporates units designed to extract R-R intervals, which are subsequently processed by LC-HHOA dense training networks to enhance the prediction accuracy of heart diseases. The model employs 152 hidden nodes, a momentum value of 0.02, and is trained over 130 epochs for optimal performance.

| Steps | Algorithm-1 // Pseudo Code for the Proposed Model |
|---|---|
| 01 | Input : Bias weight, concealed units, Epochs, Learning Rate |
| 02 | Target : Prediction of Normal/Attacks |
| 03 | Bias weight, concealed units, epochs, and learning rate should be allocated at random. |
| 04 | Set the three parameters |
| 05 | UtilizeWhile loop for true |
| 06 | Utilize equation (11) to determine the output from Classification networks |
| 07 | Utilize the formula (18) for determining the fitness function |
| 08 | Commence the For loop from t=1 to Max. iteration |
| 09 | Utilize equations (17) to allocate the bias weights & input layers |
| 10 | Utilize equation (18) for calculating the fitness function |
| 11 | Check If condition for (Fitness function is equal to threshold ) |
| 12 | jump to Step 17 |
| 13 | Otherwise |
| 14 | jump to Step 8 |
| 15 | Stop |
| 16 | Stop |
| 17 | Calculate the Output function from Dense layers using Equation (11) |
| 18 | If Output function is equal to 1 |
| 19 | //Normal is Detected |
| 20 | Else if (output function<=2&&>1) |
| 21 | // Depression is Detected |
| 22 | jump to Step 9 |
| 23 | Stop |
| 24 | Stop |
| 25 | Stop |

## 4. Results and Discussions

This section outlines the procedural aspects and evaluation outcomes derived from two separate experiments. Comprehensive ablation studies and statistical validations were rigorously performed. Ultimately, the findings are examined and juxtaposed with existing approaches, emphasizing the distinctiveness of the proposed model in facilitating future dimensionality reduction.

### 4.1 Implementation

The entire algorithm was implemented on an Intel workstation equipped with an i7 processor, an NVIDIA graphics card, 16GB of memory, and a 3.2 GHz clock speed.

### 4.2 Performance Evaluation

The recommended approach's performance was examined by utilizing metrics such as accuracy, precision, recall, specificity, and F1-score, which were subsequently examined with state-of-the-art DL approaches in the context of Fog computing. These evaluations highlight the model's advantages. Additionally, latency and computational overhead were analyzed to showcase the efficiency and low resource consumption of the proposed approach. Table 4 outlines the mathematical expressions used to calculate the performance metrics. To address issues of overfitting and ensure robust generalization, the early stopping method was implemented. This approach halts the training

process when no further improvement is observed in the validation performance over successive iterations.

***Table 4.*** *Performance measures employed in the evaluation*

| Performance Measures | Expression |
|---|---|
| Accuracy | $\frac{TP+TN}{TP+TN+FP+FN}$ |
| Recall | $\frac{TP}{TP+FN}$ x100 |
| Specificity | $\frac{TN}{TN+FP}$ |
| Precision | $\frac{TN}{TP+FP}$ |
| F1-Score | $2.\frac{Precison*Recall}{Precision+Recall}$ |

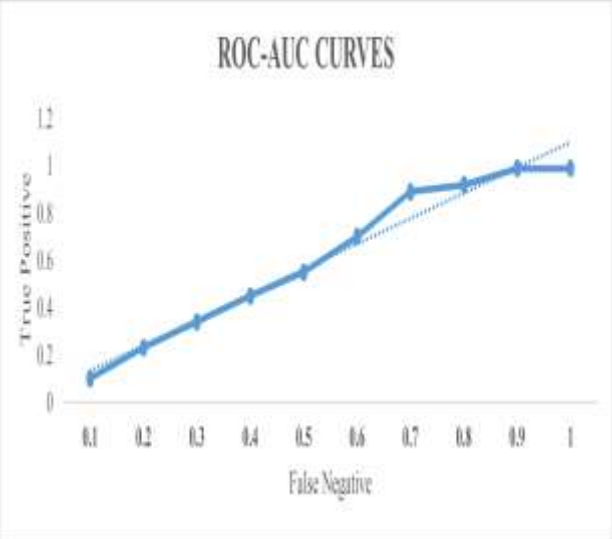


***Figure 6.*** *ROC curves for the Recommended approach In Detecting the Depression and Normal State.*

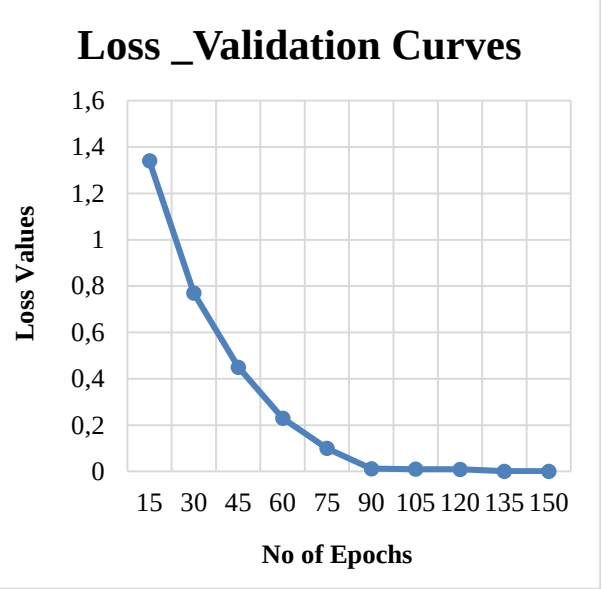


***Figure 7.*** *Loss_Validation Curves for the Recommended approach*

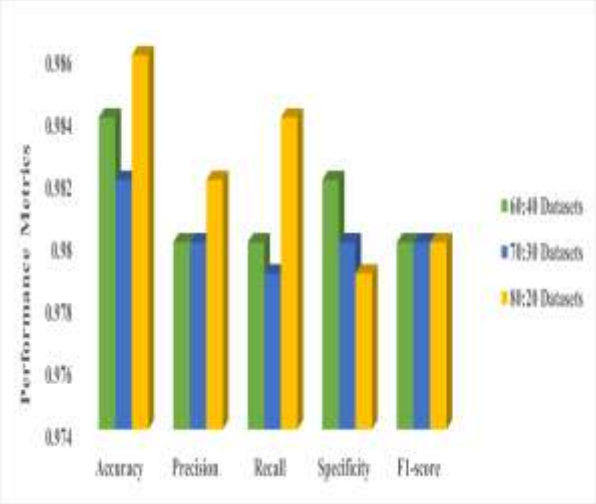


***Figure 8.*** *Performance of the Recommended approach in Detecting the Individuals' Depressed State*

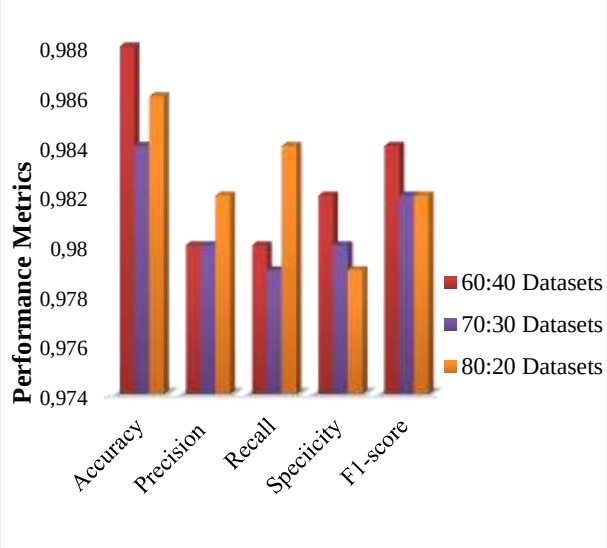


***Figure 9.*** *Performance of the Recommended approach in Detecting the Individuals' Normal State*

TP & TN are True Positive & negative, FP & FN are False Positive& negative. Prediction cases can be categorized into four distinct types: The first is TP, where the predictions are accurate, and the actual outcomes confirm their correctness. The second type is FP, which occurs when the results are incorrectly classified as positive despite being false in reality. The third is FN, where genuine instances are mistakenly labelled as negative. Finally, TN refers to scenarios where negative predictions align accurately with the actual negative outcomes. Figure 6 shows the Area-Under –Curve (AUC) –ROC (Receiver Operating Characteristics) of the Recommended approach in recognizing the depression and non –depression states. From the Figure 6, AUC is found to be 0.95 for achieving the best classification performance. The loss validation characteristics of the recommended approach is depicted in Figure 7 in which the loss is reduced for

the increased number of epochs. Hence the model has been trained with the less loss to achieve the better performance. Figure 8 and Figure 9 shows the performance of the recommended approach in two varied modes with the variations in the number of datasets. From the Figure 8 and 9, it is very clear that

***Table 5.*** *Performance of the Recommended approach in Detecting the Depressed Mode*

| Algorithms | Performance metrics | | | | |
|---|---|---|---|---|---|
| | Accuracy | Precision | Recall | Specificity | F1-score |
| RNN | 0.76 | 0.743 | 0.734 | 0.734 | 0.754 |
| LSTM | 0.80 | 0.82 | 0.812 | 0.802 | 0.80 |
| LSTM-RNN | 0.834 | 0.822 | 0.821 | 0.82 | 0.82 |
| CNN-BILSTM | 0.945 | 0.935 | 0.930 | 0.920 | 0.930 |
| BILSTM | 0.82 | 0.811 | 0.810 | 0.81 | 0.810 |
| MDAN | 0.90 | 0.89 | 0.88 | 0.874 | 0.867 |
| CNN-GRU | 0.89 | 0.88 | 0.87 | 0.86 | 0.856 |
| **Proposed Model** | **0.98** | **0.972** | **0.98** | **0.98** | **0.987** |

***Table 6.*** *Performance of the Recommended approach in Detecting the Normal Mode*

| Algorithms | Performance metrics | | | | |
|---|---|---|---|---|---|
| | Accuracy | Precision | Recall | Specificity | F1-score |
| RNN | 0.76 | 0.743 | 0.734 | 0.734 | 0.754 |
| LSTM | 0.80 | 0.82 | 0.812 | 0.802 | 0.80 |
| LSTM-RNN | 0.834 | 0.822 | 0.821 | 0.82 | 0.82 |
| CNN-BILSTM | 0.945 | 0.935 | 0.930 | 0.920 | 0.930 |
| BILSTM | 0.82 | 0.811 | 0.810 | 0.81 | 0.810 |
| MDAN | 0.90 | 0.89 | 0.88 | 0.874 | 0.867 |
| CNN-GRU | 0.89 | 0.88 | 0.87 | 0.86 | 0.856 |
| **Proposed Model** | **0.984** | **0.975** | **0.97** | **0.98** | **0.985** |

the performance of the recommended approach has attained the stable performance in classifying the depression with the variations in the datasets.

### 4.3 Performance Analysis- A Comparative View

To highlight the superiority of the recommended approach, different existing models has been considered and experimented with the input datasets which has been utilized to examine the recommended approach. Table 5 presents the performance of the recommended approach in examined with the varied residing models. Table 6 presents the performance of the varied approaches in classifying the depression and normal mode. From the table 6, RNN has yielded the least performance in recognizing the depression and normal state. CNN-GRU and CNN-BiLSTM has produced the better performance than the varied residing approach. But the Recommended approach has yielded the best performance in detecting the depression and normal state. The proposed model has produced the highest accuracy of 0.984, precision of 0.975, recall of 0.97, specificity of 0.98 and F1-score of 0.985 respectively. The integration of the Adaptive dilated convolutional networks, BIGRU with the optimized learning models has evoked the performance of the Recommended approach to achieve the better performance and stands a tall point in clinical process treatment process.

### 4.4 Statistical Analysis

To prove the excellence of the Recommended approach, several meta-heuristic algorithm incorporated with the CNN+GRU has been experimented in which the outcomes are in-cultivated and compared with the Recommended approach.The meta-heuristic algorithms employed in the experimental analysis include Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Whale Optimization Algorithm (WOA) , Spotted Hyena Optimization (SHO), Spider Optimization Algorithm (SOA), Hybrid-Reptile Search Optimization (HRSO), and White Shark Optimizers. These high-performing optimization techniques are integrated into the recommended DL frameworks, and their effectiveness is evaluated by comparing them with Shark Smell Optimizers in the model. Tables 7 and 8 showcase the results derived from various optimizer combinations in the proposed networks. Based on these outcomes, it is apparent that the recommended approach outperforms other optimization algorithms. The stability of the approach is demonstrated in Figure 9, where it is evident that the recommended approach achieves optimal performance compared to the residing approaches.

## 5. Conclusion and Future Direction

In this research work, recommended system for an automatic detection of depression using deep learning model. The integration of the dilated convolutional networks with the Bi-GRU networks with the optimized learning model is constructed for an effective combination of the depression.

***Table 7.*** *Fitness Function based Outcomes for the varied combinations of Optimizers*

| Algorithm | Best | Worst | Mean | Median | SD | Variance |
|---|---|---|---|---|---|---|
| PSO | 0.7480 | 0.625 | 0.71210 | 0.021419 | 0.06450 | $7.5x10^{-6}$ |
| GA | 0.730330 | 0.63525 | 0.69034 | 0.020202 | 0.07033 | $6.390x10^{-6}$ |
| WOA | 0.7523 | 0.6763 | 0.65372 | 0.027820 | 0.068903 | $5.892x10^{-5}$ |
| SOA | 0.77435 | 0.6902 | 0.71239 | 0.02039 | 0.054637 | $4.1290x10^{-4}$ |
| SSO | 0.80234 | 0.64389 | 0.73402 | 0.0302-2 | 0.059034 | $3.903x10^{-4}$ |
| HRSO | 0.85634 | 0.69034 | 0.74022 | 0.039303 | 0.045890 | $3.6790\ x10^{-4}$ |
| **Proposed Model** | **0.99763** | **0.81202** | **0.85640** | **0.06344** | **0.037630** | **$2.28930\ x10^{-4}$** |

***Table 8.*** *Indicator Function based Outcomes for the varied combinations of Optimizers*

| Algorithm | Best | Worst | Mean | Median | SD | Variance |
|---|---|---|---|---|---|---|
| **PSO** | 0.06453 | 0.00424 | 0.006580 | 0.006690 | 0.0001402 | $7.417\ x10^{-7}$ |
| **GA** | 0.07034 | 0.00567 | 0.006993 | 0.006890 | 0.0002302 | $6.67\ x10^{-6}$ |
| **WOA** | 0.07130 | 0.005834 | 0.007093 | 0.006903 | 0.0002789 | $5.893\ x10^{-5}$ |
| | 0.07353 | 0.005059 | 0.0070234 | 0.007045 | 0.0002890 | $5.003\ x10^{-5}$ |
| **SSO** | 0.07324 | 0.0048934 | 0.007008 | 0.0070213 | 0.000303 | $4.020\ x10^{-5}$ |
| **HRSO** | 0.06702 | 0.0045363 | 0.007102 | 0.0070231 | 0.0003450 | $3.892\ x10^{-4}$ |
| **Proposed Model** | **0.070302** | **0.0049083** | **0.00739** | **0.007240** | **0.0003893** | **$2.239\ x10^{-4}$** |

The model has been tested with the bi –modal inputs such as text inputs and audio files hence the proposed models handle the multi-modal inputs. The experimental validation is conducted using the DAIC-WIX datasets and performance metrics such as accuracy, precision, recall, specificity and F1-score are validated. Results shows that the recommended approach has shown the accuracy of 0.98, precision of 0.972, recall of 0.98, specificity of 0.98 and F1-score of 0.987 respectively. Experimental validation has also showed the recommended approaches outperformed the varied approached and integration of dwarf mongoose based optimized learning framework has improvised the performance of detection when compared with the other models. Furthermore, proposed model has statically proved to prove the role of DMO model in the improving the classification performance of the recommended approach. As the future direction, multi-modal inputs such video, audio, speech, activities can also take as the input modalities to elevate the classification performance of the recommended approach. Further, the recommended approach should be expanded for the multi-modal real time datasets and should improvise the standard of living by moving towards depression free society.

## Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## References

[1] Ramirez-Cifuentes, D., Largeron, C., Tissier, J., Baeza-Yates, R., & Freire, A. (2021). Enhanced word embedding variations for the detection of substance abuse and mental health issues on social media writings. *IEEE Access*, *9*, 130449–130471. https://doi.org/10.1109/ACCESS.2021.3112102

[2] Lam, G., Dongyan, H., Lin, W., & City, S. (2021). Context; Multi-Modal (pp. 3946–3950)..

[3] Rao, G., Zhang, Y., Zhang, L., Cong, Q., & Feng, Z. (2020). MGL-CNN: A hierarchical post representations model for identifying depressed individuals in online forums. *IEEE Access*, *8*, 32395–32403. https://doi.org/10.1109/ACCESS.2020.2973737

[4] Parapar, J., Martín Rodilla, P., Losada, D., & Crestani, F. (2023). Overview of eRisk 2023: Early risk prediction on the internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 14th International Conference of the CLEF Association, CLEF 2023*. Springer International Publishing.

[5] Zhang, Yipeng&Lyu, Hanjia& Liu, Yubao& Zhang, Xiyang& Wang, Yu &Luo, Jiebo. (2020). Monitoring Depression Trend on Twitter during the COVID-19 Pandemic: Observational Study

(Preprint). *JMIR Formative Research.* 1. 10.2196/26769.

[6] Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2021). MentalBERT: Publicly available pretrained language models for mental healthcare. *arXiv preprint arXiv:2110.15621*.

[7] Tian, H., et al. (2023). Deep learning for depression recognition from speech. *Mobile Networks and Applications*. https://doi.org/10.1007/s11036-022-02086-3

[8] Bhuvaneswari, M., & Prabha, V. L. (2023). A deep learning approach for depression detection of social media data with hybrid feature selection and attention mechanism. *Expert Systems, 40*(9), Article e13371. https://doi.org/10.1111/exsy.13371

[9] Kalpana, P., Narayana, P., Smitha, M., Dasari, K., Smerat, A., & Akram, M. (2025). Health-Fots: A latency-aware fog-based IoT environment and efficient monitoring of body's vital parameters in smart healthcare environment. *Journal of Intelligent Systems and Internet of Things*, *15*(1), 144-156. https://doi.org/10.54216/JISIoT.150112

[10] Ramirez-Cifuentes, D., Largeron, C., Tissier, J., Baeza-Yates, R., & Freire, A. (2021). Enhanced word embedding variations for the detection of substance abuse and mental health issues on social media writings. *IEEE Access, 9*, 130449–130471. https://doi.org/10.1109/ACCESS.2021.3112102

[11] Alghamdi, N. S., Hosni Mahmoud, H. A., Abraham, A., Alanazi, S. A., & García-Hernández, L. (2020). Predicting depression symptoms in an Arabic psychological forum. *IEEE Access, 8*, 57317–57334. https://doi.org/10.1109/ACCESS.2020.2981834

[12] Alhanai, T., Ghassemi, M., & Glass, J. (2018). Detecting depression with audio/text sequence modeling of interviews. *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 1716–1720. https://doi.org/10.21437/Interspeech.2018-2522

[13] Lin, L., Chen, X., Shen, Y., & Zhang, L. (2020). Towards automatic depression detection: A BiLSTM/1D CNN-based model. *Applied Sciences, 10*(23), 1–20. https://doi.org/10.3390/app10238701

[14] Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, *3*(1), 1–20. https://doi.org/10.1007/s42979-021-00958-1

[15] Beniwal, R., & Saraswat, P. (2024). A hybrid BERT-CNN approach for depression detection on social media using multimodal data. *The Computer Journal*, *67*(7), 2453–2472. https://doi.org/10.1093/comjnl/bxae018

[16] Vara Sree Yenugutalaa, N. (2024). Depression detection using machine learning and deep learning techniques. *International Journal of Research Publication and Reviews, 5*(1), 25–33.

[17] Tejaswini, V., Sathya Babu, K., & Sahoo, B. (2024). Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(1), Article 4. https://doi.org/10.1145/3569580

[18] Vandana, Marriwala, N., & Chaudhary, D. (2023). A hybrid model for depression detection using deep learning. *Measurement: Sensors, 25*, 100587. https://doi.org/10.1016/j.measen.2022.100587

[19] Khafaga, D. S., Auvdaiappan, M., Deepa, K., Abouhawwash, M., & Karim, F. K. (2023). Deep learning for depression detection using Twitter data. *Intelligent Automation & Soft Computing, 36*(2), 1301-1313. https://doi.org/10.32604/iasc.2023.033360

[20] Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep learning for depression detection from textual data. *Electronics*, *11*(5), 676. https://doi.org/10.3390/electronics11050676

[21] Kalpana, P., Kodati, S., Smitha, L., Sreekanth, D., Smerat, N., & Akram, A. (2025). Explainable AI-driven gait analysis using wearable IoT and human activity recognition. *Journal of Intelligent Systems and Internet of Things*, *15*(2), 55–75. https://doi.org/10.54216/JISIoT.150205

[22] Nadeem, A., Naveed, M., Islam Satti, M., Afzal, H., Ahmad, T., & Kim, K. I. (2022). Depression detection based on hybrid deep learning SSCL framework using self-attention mechanism: An application to social networking data. *Sensors (Basel, Switzerland)*, *22*(24), 9775. https://doi.org/10.3390/s22249775

[23] Kalpana, P., Almusawi, M., Chanti, Y., Sunil Kumar, V., & Varaprasad Rao, M. (2024). A deep reinforcement learning-based task offloading framework for edge-cloud computing. *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, Raichur, India. https://doi.org/10.1109/ICICACS60521.2024.10498232

[24] Kaggle. (2023). Kaggle: Your Machine Learning and Data Science Community. Retrieved July 3, 2023, from https://www.kaggle.com.

[25] Chung, Junyoung & Gulcehre, Caglar & Cho, KyungHyun & Bengio, Y.. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.