International Journal of Theoretical and Applied Computational Intelligence
https://ijtaci.com

Research Article

Enhancing Depression Detection through 1D Convolutional Neural Networks on DAIC-WOZ Dataset for Investigation of Visual Cues

Aayushi Chaudhari[1], Deep Kothadiya[1*], Harshit Ajakiya[1], Andrew Augustine Babu[1], and Masoumeh Soleimani[2]

[1]U and P U. Patel Department of Computer Engineering, Chandubhai S Patel Institute of Technology (CSPIT), CHARUSAT Campus, Charotar University of Science and Technology (CHARUSAT), Changa, India

[2]Department of Industrial Engineering, Clemson University, SC, United States

[*]Corresponding Author: Deep Kothadiya. Email: deepkothadiya.ce@charusat.ac.in

https://orcid.org/0000-0002-5997-1720

Received: 19/7/2025; Accepted: 13/9/2025; Published: 22/9/2025

https://doi.org/10.65278/IJTACI.2025.6

Abstract: With the increasing prevalence of depression globally, there is a growing demand for advanced, standardized diagnostic tools that can assist in early identification and intervention. This work utilises deep learning algorithms to address the growing demand for standardised and reliable diagnostic tools in depression identification. In particular, we investigate how 1D Convolutional Neural Networks (CNNs) can be utilised to analyse visual characteristics from the extensive DAIC-WOZ dataset, a collection of clinical interview sessions. The proposed architecture utilises a Deep CNN designed to discern intricate patterns in voice acoustics and facial expressions, aiming to achieve state-of-the-art precision and effectiveness. To improve the model's performance, we tested several dropout rates (0.3 and 0.5) and learning rate (0.001 and 0.0001) setups. The setup with a learning rate of 0.0001 and a dropout rate of 0.5 had the best overall performance, according to the results, with a ROC AUC of 0.79 and macro-average precision, recall, and F1-score of 0.79 across classes. According to this ideal setup, a higher dropout rate combined with a lower learning rate improves the model's ability to generalize, most likely by avoiding overfitting and enabling it to more successfully identify minute patterns in the data.

Keywords: Depression; DAIC-WOZ; CNN; Visual cues; DAIC-WOZ Dataset.

1. Introduction

Depression, a pervasive mental health disorder affecting millions globally, has traditionally been diagnosed through clinician-conducted interviews, subjective assessments, and self-reported symptoms. Gold standard tools like the Hamilton Rating Scale for Depression (HAMD) provide frameworks for evaluating symptom severity [1]. However, the reliance on subjective interpretation and variability in clinician expertise highlights the need for more objective and standardized diagnostic approaches. In response, researchers have increasingly turned to technological innovations, particularly Automatic Depression Estimation (ADE) systems [2]. ADE systems leverage advancements in machine learning, speech recognition, and computer vision to analyze audiovisual cues for estimating depression severity [3]. Deep learning techniques, especially convolutional neural networks (CNNs), have gained popularity recently because they provide improved feature extraction efficiency and accuracy [4]. CNNs, with their ability to autonomously learn discriminative features from raw data, have emerged as powerful tools in ADE systems, surpassing traditional handcrafted features like Local Binary Patterns (LBP) and Facial Action Units (FAUs) [5].

The use of CNNs in depression detection is particularly promising when analysing complex, multimodal data sources, such as audio and video recordings. The DAIC-WOZ dataset, a comprehensive corpus containing clinical interview sessions with audio, video, and transcript data, provides an ideal foundation for developing and testing CNN-based depression detection models [6]. This dataset includes detailed facial feature points, action unit detections, and head pose data, as well as audio features and interview transcripts, enabling a multi-faceted approach to analyzing signs of depression [7].

In this study, we proposed a deep learning-based architecture Deep CNNs for extraction of facial features to detect expressions. The Deep CNN is designed to interpret speech acoustic characteristics and visual information from facial expressions. The proposed architecture includes four stages of convolution with increasing filter sizes, batch normalization, max pooling, and dense layers with dropout for regularization, leading to a binary classification output [8]. Techniques like late fusion, which combines multiple modalities at a later stage of the model pipeline, are employed to enhance system robustness and generalizability [9]. This work intends to add to the expanding corpus of literature on automated depression diagnosis by emphasizing the potential of deep learning techniques to offer more impartial and scalable diagnostic instruments for mental health treatment. By leveraging the rich, multimodal data provided by the DAIC-WOZ dataset and optimizing CNN architectures, we seek to advance the state-of-the-art in depression detection.

The rest of the article is organized as: Section 2 discussed on recent research work on emotion recognition with deep learning or machine learning. Section 3 discuss proposed methodology of study, which involve deep 1D Convolution for feature extraction. Section 4 discuss results and comparative study with SOTA deep learning models for emotion recognition.

2. Literature Review

Gold-standard assessments like the Hamilton Rating Scale for Depression (HAMD) have provided a framework for evaluating symptom severity [13]. However, the reliance on subjective interpretation and the variability in clinician expertise have highlighted the need for more objective and standardized approaches. In response to the limitations of traditional diagnostic approaches, researchers have turned to technological innovations, particularly Automatic Depression Estimation (ADE) systems. Leveraging techniques from machine learning, speech recognition, and computer vision, these systems analyze

audiovisual cues to estimate depression severity [14]. Notably, recent years have witnessed a shift towards deep learning approaches, which offer enhanced accuracy and efficiency in feature extraction [15].

L. He et al., 2020 [16] provides a detailed review of approaches to depression detection using clinician-conducted interviews, subjective assessments, and self-reported symptoms. Deep Convolutional Neural Networks (DCNN) have emerged as powerful tools for extracting multi-scale feature representations in ADE systems. Unlike traditional hand-crafted features such as Local Binary Patterns (LBP) and Facial Action Units (FAUs), deep learning models can autonomously learn discriminative features from raw data, leading to improved performance in depression recognition tasks [17]. However, challenges remain in optimizing network architectures and integrating multimodal data sources to enhance ADE system robustness and generalizability. In recent years, there has been a growing interest in the developing automated systems for detecting depression, particularly through the analysis of audio signals containing speech segments. This interest stems from the recognition of the rich latent information embedded within speech, which can provide valuable insights into an individual's mental state [18]. However, many existing approaches predominantly rely on stacking deep neural networks, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), which may not adequately capture the diverse range of depression-related features present in speech. Moreover, these single-stream networks often entail significant computational overhead and memory requirements.

To address these limitations, Yin [19] introduces a novel approach based on a parallel convolutional neural network (CNN) and transformer model. This model aims to extract effective representations of depression from speech signals while maintaining computational efficiency. Specifically, low-level mel-frequency cepstral coefficient (MFCC) features are utilized as input, with a parallel CNN module employed to capture local information. Additionally, a transformer module with an improved linear attention mechanism is incorporated to capture temporal sequential information in speech. Unlike conventional RNN structures, the proposed transformer leverages linear attention mechanisms with kernel functions to reduce computational complexity. Morales [20] conducted a survey aimed at bridging the subfields by reviewing depression detection systems across subfields and modalities. This paper discusses how depression has been defined and annotated in detection systems, what types of depression data exist or may be gathered for depression detection systems, and which (multimodal) indicators have been employed for automatic depression detection and their assessment metrics.

3. Methodology

3.1. Deep CNN

A Deep Convolutional Neural Network is a form of neural network that is specifically intended to analyze visual information. It is made up of several layers of convolution that use images as input to automatically and adaptively learn the spatial hierarchies of various features [21]. Deep CNNs are highly effective at tasks like object detection, semantic segmentation, and picture classification because they can capture complex patterns and features at different levels of abstraction [22]. They provide leading-edge performance in numerous difficult visual recognition tasks and are extensively employed in domains including computer vision, medical imaging, autonomous driving, and facial recognition [23].

3.2 Proposed Architecture

In this work, we offer a deep learning-based architecture that employs a 1D deep CNN for video modality. Openface has been used to extract the facial features per frame for facial landmarks, these features further form time series data in sequence, where 1D CNN is used to apply on temporal axis. 2D CNN has

impressive processing capabilities for spatial data like images with 2D filters (e.g., 3x3), the 1D CNN uses filters of size 3 to capture temporal dependencies across consecutive time steps in the sequential data, giving impressive results for time series data. Figure 1 represents the detailed architecture of proposed work.

Deep Convolutional Neural Networks (Deep CNNs) are used in proposed design to interpret speech acoustic characteristics and visual information from facial expressions. The acceptable dimension of the input layer is (12447, 382). To reduce dimensionality while retaining key features, it has four stages of convolution with 32, 64, 128, and 256 filters, respectively. This progression allows the model to learn increasingly abstract and complex features at each stage. The early layers capture basic patterns, while the deeper layers identify higher-level features, essential for depression detection in time-series data [24]. Batch normalization and max pooling are applied after each convolutional layer. After flattening the output, the network has four dense layers of 512, 256, 128, and 64 neurons. Dropout is applied for regularization by randomly deactivating a portion of neurons during training, which helps prevent overfitting and encourages the network to learn more robust features [25]. Algorithm 1 present pseudocode for proposed architecture having phase 01 as feature extraction and phase 02 as classification.

Algorithm 01

1: Initialize Parameter = $\alpha = [\ldots\ldots]$
// Feature Extraction
2: for i in 1 to length(filters) do
3: conv_output ← Convolution1D(input_layer, filters[i], kernel_size)
4: conv_output ← BatchNormalization(conv_output)
5: conv_output ← MaxPooling1D(conv_output)
6: input_layer ← conv_output
7: end for
8: lattened_output ← Flatten(input_layer)
// Classification
9: for j in 1 to length(dense_neurons) do
10: dense_output ← Dense(flattened_output, dense_neurons[j])
11: dense_output ← Dropout(dense_output, dropout_rate)
12: flattened_output ← dense_output
13: end for
14: output ← Dense(flattened_output, 1, activation=”Softmax”)

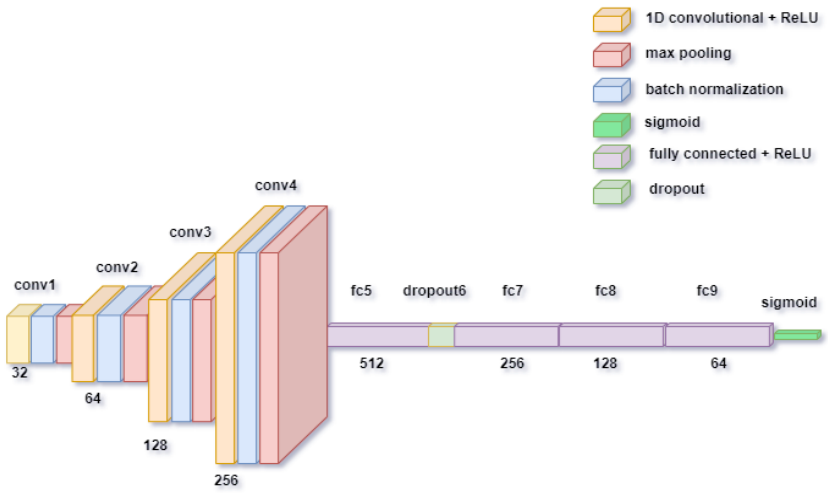

Figure 1: Architecture for 1D Convolutional Neural Network

The last layer of a fully connected neural network results in a binary classification output with the help of a sigmoid activation function. We have used the Adam optimizer with varying learning rates to gather the result [26]. The model utilizes the binary cross-entropy loss function, which is well-suited for binary classification tasks [27].

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^{x}}{1+e^{x}} = 1 - \sigma(-x) \quad (1)$$

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \,.\, log(p(y_i)) + (1-y_i)\,.\, log(1-p(y_i)) \quad (2)$$

4. Experimental result and analysis

4.1 Dataset

For this study, the Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ) dataset, a subset of the broader Distress Analysis Interview Corpus (DAIC), is utilised. This corpus contains 189 recordings of clinical interview sessions ranging from 300 to 492 conducted with Ellie, the virtual interviewer, administered by a human interviewer in another room [10]. To aid in the identification of mental health issues like anxiety, depression, and post-traumatic stress disorder, the interview format was created. For every session, the dataset contains the following information,

Audio and video recordings of the interviews ranged from 7 to 33 minutes (average 16 minutes), as well as transcripts of the interview conversations.

Each participant's folder consists of these files:

- XXX_CLNF_features.txt: Contains 68 2D feature points on the face.
- XXX_CLNF_AUs.csv: Provides frame-level action unit detections with regression values and binary presence flags for 19 facial action units.
- XXX_CLNF_features3D.txt: Includes 68 3D facial landmark points per frame, with coordinates in millimetres in the world/camera coordinate space.
- XXX_CLNF_gaze.txt: Contains frame-level eye gaze direction vectors, provided in both world and head coordinate spaces.
- XXX_CLNF_hog.bin: Stores histograms of oriented gradients (HOG) features extracted from the aligned face region as a 4464-dimensional vector per frame.

- XXX_CLNF_pose.txt: Provides the 3D head position (X, Y, Z) and rotation (Rx, Ry, Rz) per frame in world coordinates and Euler angles, respectively.
- XXX_AUDIO.wav: The 16kHz audio recordings from a head-mounted micro-phone, with identifiable utterances scrubbed by setting the waveform to zero. These scrubs can be tracked using the transcript files and the "scrubbed_entry" keyword. Some bleed-over from the virtual interviewer may be present.
- XXX_TRANSCRIPT.csv: Contains the transcripts of the interviews, following specific conventions such as marking incomplete words, overlapping speech timestamps, and including automated transcriptions for participants after Partici-pant 363 with unique utterance identifiers. A detailed transcription manual is provided.
- XXX_COVAREP.csv: Contains low-level audio features extracted every 10 ms (100 Hz), including F0, voice/unvoiced flag, NAQ, QOQ, H1H2, PSP, MDQ, peakSlope, Rd, Rd_conf, MCEP, and HMPD coefficients. Unvoiced regions have unreliable values for some features, and scrubbed entries are set to zero.
- XXX_FORMANT.csv: Includes the trajectories of the first 5 formant frequencies, with scrubbed entries again set to zero values.

4.2 Results analysis

In this section, we present the results obtained from the proposed deep learning-based architecture using the DAIC-WOZ dataset. We evaluated the performance of the model by varying two key hyperparameters: dropout rate and learning rate, two hyperparameters are used in this model. This led to four variants (2 x 2) where their performances were evaluated using classification measures such as accuracy, precision, recall, F1-score and ROC AUC [28].

The model's performance was evaluated under the following hyperparameter configurations: Configuration 1 (Dropout: 0.2, Learning Rate: 0.001), Configuration 2 (Dropout: 0.2, Learning Rate: 0.0001), Configuration 3 (Dropout: 0.5, Learning Rate: 0.001), Configuration 4 (Dropout: 0.5, Learning Rate: 0.0001). For each configuration, the model was trained and tested on the dataset, and the classification reports and ROC AUC curves were analyzed to compare performance across configurations. The Adam optimizer was used to update the model weights, which is well-suited for this task due to its efficiency in handling sparse gradients and its adaptive learning rate [29].

The training data exhibits a class imbalance, with 77 participants labelled as not depressed and only 30 as depressed. Figure 2 represents a graphical view of depressed and non-depressed participants.

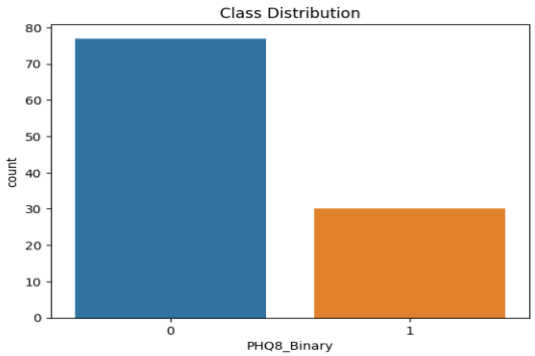

Figure 2: Class distribution for depressed and non-depressed participants

The dataset has two gender labels, with 0 representing female and 1 representing male. In their study, Bailey and Plumbley [11] noted that the dataset's gender distribution is skewed, which may introduce

gender bias in depression detection models that primarily rely on audio features. Figure 3 represents class distribution and gender representation for depressed and non-depressed participants.

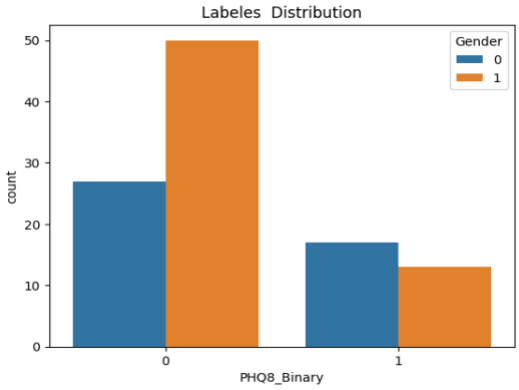

Figure 3: Class distribution for depressed and non-depressed participants along with their gender representation

The validation set contains (23, 12) not-depressed and depressed labels, respectively, while the test set has (33, 14) for not-depressed and depressed labels.

Note on Depression Labeling: The Patient Health Questionnaire-8 (PHQ-8) is used to determine depression labels in the DAIC-WOZ dataset. Eight typical symptoms of depression, including hopelessness, loss of interest in activities, and changes in appetite, are evaluated by the PHQ-8, a short questionnaire. Higher scores indicate more severe depression; values range from 0 to 24. A positive screen for depression indicates a probability of depressive disorder, with a score of 10 or above [12].

4.3 Performance metrics

The results of the experiments are summarized in Table 1 below which outlines the detailed performance of each configuration. The Class column is used to show Metrics separately for classes 0 and 1, along with the macro-average score that reflects the overall performance across both classes. Figure 4 represents the confusion matrix for representing results of depressed and non-depressed participants for dropout value 0.5. Figure 5 represents the ROC curve of the result at dropout value 0.5.

Table 1: Result comparison by applying different dropout instances

Dropout	Learning rate	Class	Precision	Recall	F1-Score	ROC curve area
0.3	0.001	0	0.33	0.91	0.49	0.64
		1	0.94	0.44	0.6	
		Macro avg	0.64	0.68	0.55	
0.5	0.001	0	0.8	0.83	0.81	0.75
		1	0.71	0.67	0.69	
		Macro avg	0.75	0.75	0.75	
0.3	0.0001	0	0.1	1	0.18	0.55
		1	1	0.39	0.56	
		Macro avg	0.55	0.69	0.37	
0.5	**0.0001**	**0**	**0.87**	**0.84**	**0.85**	**0.79**
		1	**0.71**	**0.75**	**0.73**	
		Macro avg	**0.79**	**0.79**	**0.79**	

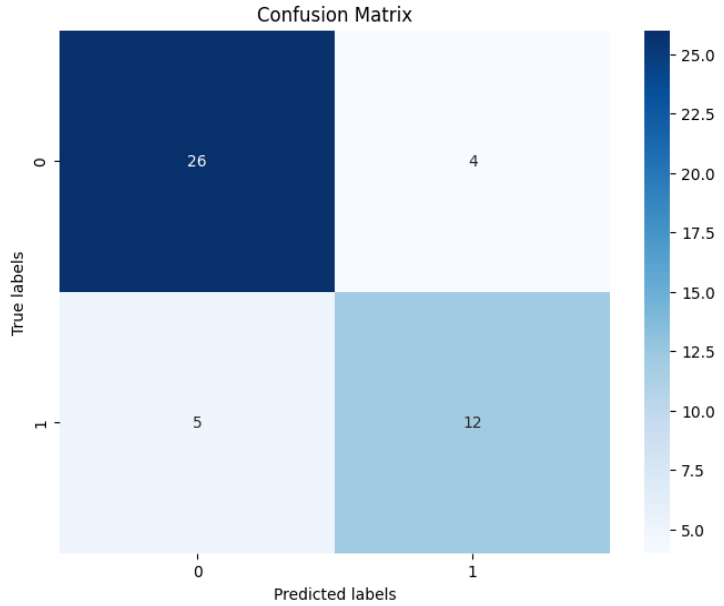

Figure 4: Confusion matrix for representing depressed and non-depressed for dropout=0.5 and lr=0.0001

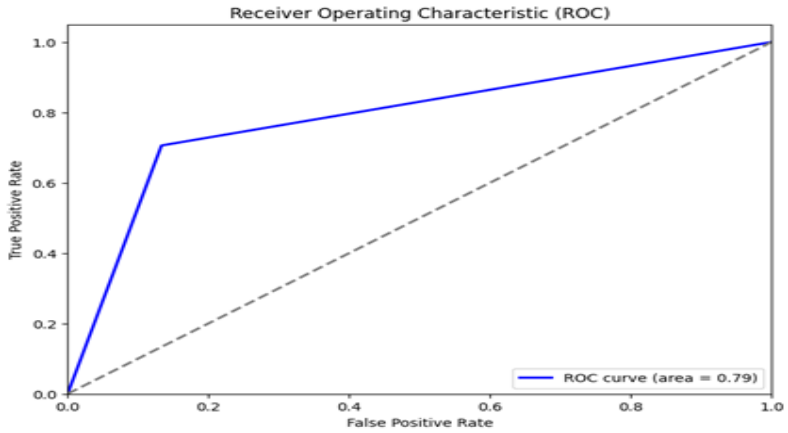

Figure 5: ROC curve for true and false positive rate, for dropout=0.5 and lr=0.0001

4.4 Discussion

The performance of the 1D CNN model on the DAIC-WOZ dataset varied significantly across different hyperparameter configurations, particularly between dropout rates (0.3 and 0.5) and learning rates (0.001 and 0.0001) [15]. Among the configurations tested, the combination of a dropout rate of 0.5 and a learning rate of 0.0001 yielded the best overall performance. This configuration achieved a macro-average precision, recall, and F1-score of 0.79 across classes, with a ROC AUC of 0.79. This suggests that a higher dropout rate coupled with a lower learning rate helped the model generalize better, likely by preventing overfitting and allowing the model to better capture subtle patterns in the data [30].

Conversely, the configuration with a dropout rate of 0.3 and a learning rate of 0.001 performed the worst, with a macro-average F1-score of 0.55 and a ROC AUC of 0.64. The high dropout rate in the best configuration may have helped the model avoid over-relying on specific features, while the lower learning rate allowed the model to converge more gradually, avoiding the pitfalls of rapid and possibly erratic updates during training [31]. Table 2 states the comparative analysis of proposed work with state-of-the-art work on the Diac-Woz dataset.

Table 2: Comparative analysis of proposed work in current literature using Diac-Woz dataset.

Model	Features used	Precision (%)	Recall (%)	F1-Score (%)
Dilated CNN on landmarks+pose [21]	2D landmarks + head-pose	0.77	0.83	0.78
Hybrid SVM–CNN– from multimodal study (video only) [22]	Facial features	0.38	0.58	0.35
LSTM – from multimodal baseline **(**video only) [22]	Facial features	0.49	0.68	0.57
Proposed Model	Facial features	0.79	0.79	0.79

5. Conclusion

This study addresses the need for more objective and standardized diagnostic tools in depression detection by leveraging deep learning approaches, specifically 1D Convolutional Neural Networks (CNNs), to analyze visual features from the DAIC-WOZ dataset. The proposed model focuses on extracting meaningful patterns from facial expressions and acoustic features, aiming to improve the accuracy and efficiency of depression detection compared to traditional methods. By evaluating various configurations of dropout rates and learning rates, our results demonstrate the significant impact of hyperparameter tuning on the performance of CNNs in detecting depression. Future research could expand on these findings by integrating additional data modalities, such as audio and text, and exploring further advancements in CNN architectures and training strategies. Ultimately, our study demonstrates that with careful optimization, deep learning models can help us transform the depression detection approach, making it more objective, accessible, and effective.

Acknowledgement: I would like to share my heartfelt gratitude to Mr. Harshit Ajakiya and Mr. Andrew Augustin for their valuable contribution to this work.

Data Availability Statement: All data generated or analyzed during this study are reported in this published article at Ref. [10].

Funding: No specific funding received for this research.

Conflicts of Interest: No conflict of interest is stated by the author.

Authors contributions. Conceptualization: AC, DK, HA; methodology: AC, DK, HA, validation: HA, AAB, MS; writing—original draft preparation, HA, AAB, MS; writing—review and editing: AC, DK, HA; visualization: AC, DK, MS; supervision: HA, AAB, MS; project administration: AC, DK, HA. All authors had approved the final version.

References

[1] Garcia-Garcia, J. M., Penichet, V. M. R., and Lozano, M. D. "Emotion detection". In Proceedings of the Xviii International Conference on Human Computer Interaction New York, NY, USA: ACM 1-8, (2017).

[2] Uddin, M. A., Joolee, J. B., and Sohn, K. (2023). "Deep multi-modal network based automated depression severity estimation". *IEEE Transactions on Affective Computing*, 14(3), 2153-2167.

[3] Jani, S., Kothadiya, D., and Chaudhari, A., "Comprehensive analysis on sentiment analysis using bidirectional sequential models", In (Eds.), *Studies in Smart Technologies*, Singapore: Springer Nature Singapore, 31-42, (2025).

[4] Kothadiya, D. R., Bhatt, C. M., and Rida, I., "Simsiam network based self-supervised model for sign language recognition", In (Eds.), *Communications in Computer and Information Science*, Cham: Springer Nature Switzerland, 3-13, (2024).

[5] Sandbach, G., Zafeiriou, S., and Pantic, M. "Binary pattern analysis for 3D facial action unit detection". In Procedings of the British Machine Vision Conference 2012: British Machine Vision Association 119.1-119.12, (2012).

[6] Squires, M., Tao, X., Elangovan, S., et al., (2023). "Deep learning and machine learning in psychiatry: A survey of current progress in depression detection, diagnosis and treatment". *Brain Informatics*, 10(1), 1-19.

[7] Pinto, S. J., and Parente, M. (2024). "Comprehensive review of depression detection techniques based on machine learning approach". *Soft Computing*, *28*(17-18), 10701-10725.

[8] Chiong, R., Budhi, G. S., Dhakal, S., et al., (2021). "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts". *Computers in Biology and Medicine*, 135, 104499.

[9] Hu, S., Zhou, H., Hergul, M., et al., (2023). "Multi 3 woz: A multilingual, multi-domain, multi-parallel dataset for training and evaluating culturally adapted task-oriented dialog systems". *Transactions of the Association for Computational Linguistics*, 11, 1396-1415.

[10] Peskin, R. R. (2017). "DAIC-WOZ: A dataset for depression analysis using interview recordings". *Lang. Soc. Psychol.*, *36*, 356–370.

[11] Bailey, A., and Plumbley, M. D. "Gender bias in depression detection using audio features". In 2021 29th European Signal Processing Conference (eusipco): IEEE 596-600, (2021).

[12] Kroenke, K., Strine, T. W., Spitzer, R. L., et al., (2009). "The phq-8 as a measure of current depression in the general population". *Journal of Affective Disorders*, *114*(1-3), 163-173.

[13] Hamilton, M., "The hamilton rating scale for depression", In (Eds.), *Assessment of Depression*, Berlin, Heidelberg: Springer Berlin Heidelberg, 143-152, (1986).

[14] Pampouchidou, A., Pediaditis, M., Kazantzaki, E., et al., (2020). "Automated facial video-based recognition of depression and anxiety symptom severity: Cross-corpus validation". *Machine Vision and Applications*, *31*(4).

[15] He, L., and Cao, C. (2018). "Automated depression analysis using convolutional neural networks from speech". *Journal of Biomedical Informatics*, *83*, 103-111.

[16] Meethongjan, K., Dzulkifli, M., Rehman, A., et al., (2013). "An intelligent fused approach for face recognition". *Journal of Intelligent Systems*, 22(2), 197-212.

[17] Kothadiya, D., Bhatt, C., Khan, A. R., et al., "Explainable ai for medical science: A comprehensive survey, current challenges, and possible directions", In (Eds.), *Explainable Artificial Intelligence in Medical Imaging*, Boca Raton: Auerbach Publications, 36-57, (2025).

[18] Alsenani, B., Esposito, A., Vinciarelli, A., et al., "Assessing privacy risks of attribute inference attacks against speech-based depression detection system", In (Eds.), *Frontiers in Artificial Intelligence and Applications*, IOS Press, (2024).

[19] Yin, F., Du, J., Xu, X., et al., (2023). "Depression detection in speech using transformer and parallel convolutional neural networks". *Electronics*, 12(2), 328.

[20] Chaudhari, A., Bhatt, C., Krishna, A., et al., (2022). "Vitfer: Facial emotion recognition with vision transformers". *Applied System Innovation*, *5*(4), 80.

[21] Shah, K., Shah, K., Chaudhari, A., et al., "Comprehensive analysis of deep learning models for brain tumor detection from medical imaging", In (Eds.), *Lecture Notes in Networks and Systems*, Singapore: Springer Nature Singapore, 339-351, (2024).

[22] Kothadiya, D., Rehman, A., AlGhofaily, B., et al., (2025). "Vgx: Vgg19-based gradient explainer interpretable architecture for brain tumor detection in microscopy magnetic resonance imaging (mmri)". *Microscopy Research and Technique*, *88*(5), 1544-1554.

[23] Yang, Y., Fairbairn, C., and Cohn, J. F. (2013). "Detecting depression severity from vocal prosody". *IEEE Transactions on Affective Computing*, *4*(2), 142-150.

[24] Kothadiya, D. R., Bhatt, C., Chaudhari, A., et al., "Gujformer: A vision transformer-based architecture for gujarati handwritten character recognition", In (Eds.), *Lecture Notes in Networks and Systems*, Singapore: Springer Nature Singapore, 89-101, (2024).

[25] Guo, Y., Zhu, C., Hao, S., et al., (2023). "Automatic depression detection via learning and fusing features from visual cues". *IEEE Transactions on Computational Social Systems*, *10*(5), 2806-2813.

[26] Gajjar, D. B., Faldu, P., Kothadiya, D. R., et al., (2025). "Devitc: Deep-vision transformer to recognize originality of currency". *Computer*, *58*(5), 48-56.

[27] Ghosh, D., Karande, H., Gite, S., et al., (2024). "Psychological disorder detection: A multimodal approach using a transformer-based hybrid model". *Methodsx*, 13, 102976.

[28] Srivastava, N., Hinton, G., Krizhevsky, A., et al., (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.

[29] Nishad, D. K., Tiwari, A. N., and Khalid, S. (2025). Multi-Objective Optimization Simulation of Unified Power Quality Conditioner (UPQC). *International Journal of Theoretical & Applied Computational Intelligence*, 2025, 71-83.

[30] Pang, M., Ting, K. M., Zhao, P., et al., (2022). "Improving deep forest by screening". *IEEE Transactions on Knowledge and Data Engineering*, *34*(9), 4298-4312.

[31] Naqi, S. A. E. A., Iqbal, K., Khan, A. A., et al., (2025). "Diseases detection from apple leaf using deep transfer learning approach". *International Journal of Theoretical & Applied Computational Intelligence*, *2025*, 57–70.