Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Design and Implementation of a Hybrid Multimodal Deep Learning Model for Mental Health Detection with Adaptive Conversational Support

Prof. Jaya Nag Mathur
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Prof. Chetana Shravage
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Anjali Sinkar
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Joya Tamboli
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Shraddha Satpute
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Girija Raskar
Dept. Artificial Intelligence & Data Science
Dr. D. Y. Patil Institute of Technology
Pimpri, India

Abstract*—It's increasingly common worldwide to see mental health issues like stress, anxiety, and depression. A lot of people dealing with these problems don't get diagnosed because they can't easily access help or because of the social stigma involved. But now, with progress in AI and technology that can recognize emotions, we can build systems that watch how users are feeling by looking at different kinds of data. One promising idea is to use this tech to spot people who might be struggling with their mental health and offer them help sooner than we usually can. This project aims to create a special kind of system called a Hybrid Multimodal Deep Learning Architecture for Detecting Mental Health and Providing Adaptive Conversational Support. This system will bring together different technologies, like analyzing text, recognizing facial expressions, and using psychological questionnaires. It will be using various ML and deep learning ways, such as Logistic Regression, SVMs, LSTM, BERT Transformers, and CNNs. The main idea is to blend these technologies using these algorithms to create a thorough way to figure out mental health indicators from what users put into the system. The text-based parts will use NLP. Facial emotion recognition will use DL models which are trained on FER2013 dataset. For structured mental health answers, it will use a questionnaire similar to the DASS-21. The information from each part will be combined using a multimodal fusion technique to figure out the overall risk level for mental health. Once a person's emotional state is known, a smart chatbot will offer support and help them reach their desired emotional state. The system also includes a way to detect suicide risk.Also a chat bot running live, to guide and detect the face and eye movements to identify different mood shifts while you are typing. The chat-bot also give some guided replies when it detects how user feels and makes the chats feel warmer, and calmer for user. If red flags such as deep despair or talk of suicide pop up then it instantly displays hotlines, emergency numbers and expert resources without any delay

***Index Terms*—Mental Health Detection, Multimodal Deep Learning, Emotion Recognition, Conversational AI, BERT, CNN, Affective Computing.**

I. Introduction

Mental health issues are very prevalent as of late worldwide - this includes mental illnesses as anxiety, stress, and depression that can also greatly hinder a person's emotional quality of life and daily functioning. The World Health Organization (WHO) stated that nearly 1 in 8 individuals worldwide is living with a mental health problem [2], [20]. Early detection and intervention are key in reducing both the severity and long-term effects of mental illnesses. However, individuals may be reluctant to seek help because of stigma, a lack of understanding about what mental health is and what professionals do, limited access to care [17], [27]. Due to recent advancements in AI and ML, there are now intelligent systems that can detect mental illness based on behavioral and emotional information. Through the use of natural language processing (NLP), these systems are able to analyze different types of text-based information (e.g.speech from conversations) to determine if there is a linguistic pattern that would be suggestive of depression or anxiety [3],[6].Affective computing techniques are able to identify the feelings of individuals (based on their facial features or actions) as demonstrated by systems with deep learning methods, such as CNNs [4], [23]. Furthermore, transformer-based architectures, like BERT, enhance an understanding of language in context, which aids in providing a more accurate determination of the emotion behind the text provided as input [11], [40]. In addition to emotional recognition abilities, AI-enabled conversational agents are well-positioned to assist individuals in managing their mental health.Today, newer therapy chatbots utilize state-of-the-art NLP methods and hybrid conversational structures to offer customized support and direction to users [12], [13], [15].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

II. Literature Review

Artificial intelligence (AI) is being increasingly employed for analyzing behavioral and emotional data by employing computer-based techniques for the detection of mental health conditions as stress, anxiety, and depression. The primary focus of the initial studies has been to use machine learning techniques for analyzing textual and structured psychological data. Researchers such as Chaturvedi et al. [1] and Kumar and Singh [3] have attempted to use supervised algorithms such as SVM and the Random Forest for the classification of depression symptoms by using datasets such as DASS-21 and PHQ-9.

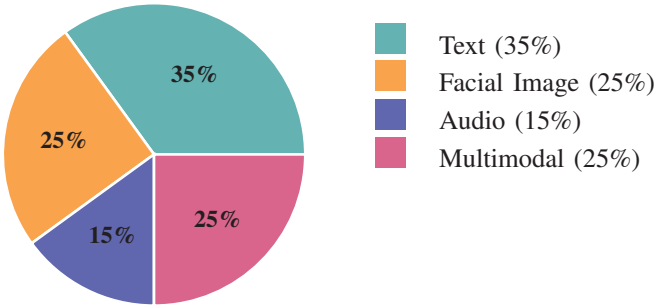

Fig. 1. Contribution of Different Modalities

These traditional machine learning models had acceptable accuracy rates, ranging from **70% to 78%**, but since they relied on human-designed features, they had trouble identifying complex emotional patterns. Because social media sites provide a wealth of textual data, researchers may now use NLP techniques to identify mental health disorders. By analyzing language patterns taken from online exchanges, Benton et al. [6] demonstrated how multi-task learning models may be used to predict psychological disorders. In the area of sentiment analysis and emotion recognition in texts, the accuracy has been improved with the introduction of **BERT** and **GPT** transformer-based architectures that have greatly improved the context understanding of texts [14]. Another important area of research in the application of emotional computing algorithms is **Facial Emotion Recognition** using DL algorithms. The accuracy of these systems has been greatly improved compared to the earlier systems that relied on the use of visual characteristics. When it comes to the recognition of emotional states from face photos, Convolutional Neural Networks (CNN) with the help of datasets like **FER-2013** and **AffectNet** have achieved an accuracy of **82%** to **87%** [7], [25], [27]. Emotional signals from speech or textual interactions have also been examined by using sequential deep learning methods like RNN and LSTM networks. When applied to data sets like Reddit or IEMOCAP, these methods have shown that temporal trends in emotional data can be recognized with an accuracy between **84%** and **88%** [10], [30]. However, recent studies have focused more on **Multimodal Emotion Recognition**, which involves the integration of different data inputs such as text, voice, and facial expressions. Tzirakis et al. in [8], [12] and Han et al. in [9] have developed a **Multimodal Framework** that utilizes deep learning techniques to blend different emotional data inputs and attain an accuracy rate of **92% to 95%**. This implies that a deeper understanding of human emotions can be achieved by integrating different data inputs.

TABLE I
Comparative Analysis of AI Approaches Used in Mental Health Detection

Approach	Data Modality	Model Type	Benchmark Dataset	Accuracy / F1 (%)
SVM / Random Forest [3], [21]	Text (survey responses)	Traditional Machine Learning	DASS-21, PHQ-9	70–78
CNN-Based Emotion Recognition [7], [25], [27]	Facial Images	Deep Learning	FER-2013, AffectNet	82–87
RNN / LSTM Models [10], [30]	Text / Audio Conversations	Sequential Deep Learning	Reddit, IEMOCAP	84–88
Transformer Models (BERT, GPT) [14], [17], [18]	Dialogue / Contextual Text	NLP Transformer Architecture	Twitter, Woebot Logs	89–93
Multimodal Fusion Systems [8], [12], [29]	**Text + Facial + Audio**	**Hybrid Deep Learning**	**AffectNet, DAIC-WOZ**	**92–95**

As mentioned earlier, recent studies have shown that there is a shift from the traditional ML techniques towards DL and multimodal techniques for mental health identification. Though traditional techniques have shown good results in mental health identification, the use of multimodal techniques for data detection has shown better results.

III. Methodology

The suggested system is anticipated to employ a Hybrid Multimodal Deep Learning Framework for Detecting Mental Health Disorders and Versatile Conversation Support. The suggested system employs text data evaluation, facial emotion evaluation, and psychological evaluation assessment for possible psychological health assessment.

A. Data Collection

The system makes use of two public data sets and real-time user data.

1) Reddit Depression Dataset: The dataset consists of textual posts made by people who are depressed as well as those who

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

are not. The dataset is used to build a natural language processing algorithm that recognizes depressed language patterns.

2) FER2013 Facial Emotion Dataset: This dataset includes seven emotional classes: angry, disgusted, afraid, joyous, neutral, sad, and startled. The dataset is then being used to build a DL model that recognizes face expressions.

3) Questionnaire-Based Input: The DASS-21, is the basis for a structured psychiatric questionnaire. Included are questions about emotional condition and stress.

B. Data Preprocessing and Classification

The user's answers and chats are handled via Natural Language Processing (NLP). The pre-processing of the data includes the following steps: Cleaning and normalizing text

Tokenization

• Handling Stop Words • Extraction of TF-IDF Features

For text classification, a number of DL and ML approaches have been used. The following algorithms are among the methods used for the project: Logistic Regression , SVMs • LSTM • The Transformer Model, or BERT

Of the algorithms used, the BERT algorithm offers the most contextual information.

C. Facial Emotion Recognition

The FER2013 dataset is used to identify face expressions using deep learning algorithms. These methods are as follows:

1) Haar Cascade Classifiers for face detection.

2) Preprocessing and normalization of images .

3) Convolutional neural networks for the extraction of features .

4) Convolutional Neural Networks and MobileNet for Emotion Classification .

D. Questionnaire-Based Psychological Screening

In order to determine the symptoms associated with anxiety, stress, and depression, the questionnaire module analyzes user replies.

The structured psychological assessment complements the machine learning algorithms' predictions.

E. Multimodal Fusion Mechanism

Weighted decision fusion is utilized in improving the reliability of the results from text prediction, results from face expression detection, and results from the questionnaire score. This is how the final mental health risk score is determined: 0.5 × Text Prediction Score 0.3 × Emotion Detection Score 0.2 × Questionnaire Score The results from mental health detection can be improved by combining multiple emotional signals at any given time.

F. Adaptive Conversational Chatbot

Based on the identified level of mental health risk, an AI-powered conversational chatbot engages with the user. It offers encouraging reactions and emotional support, based on detected emotional state of a user. The chatbot uses a number of combination :

• Rules-based conversational responses

• Natural language processing for intent detection

The adaptive conversational aid helps users communicate their emotions.

G. Report Generation

Following this evaluation procedure, the system creates a mental health screening report based on the following:

• An overview of the questionnaire replies

• A facial expression of emotion .

• Risk level for mental health .

• Interventions or actions to be performed .

A user's mental health state is summarized in the mental health screening report.

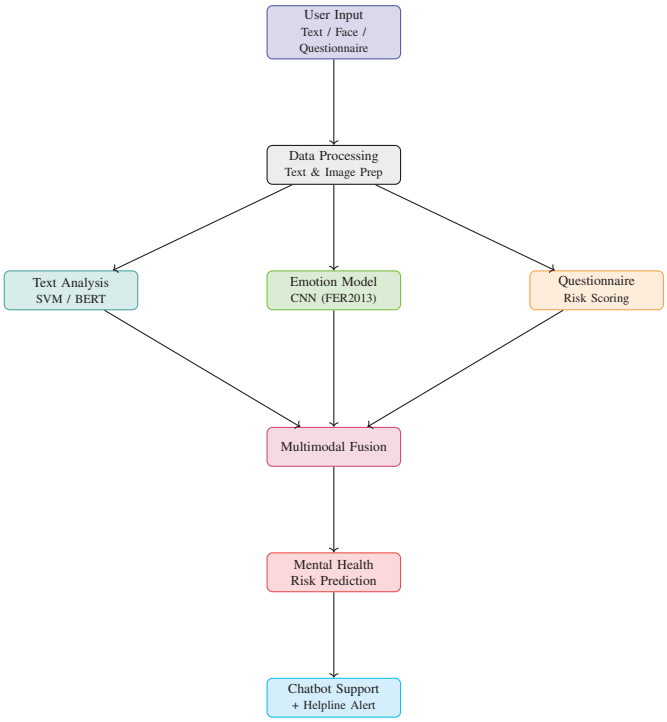

Fig. 2. Proposed Multimodal Mental Health Detection Framework

IV. Results and Discussion

The artificial intelligence techniques used for mental health detection are included in the experimental review. These techniques include machine learning, deep learning, and multimodal fusion. The effectiveness of these techniques is evaluated using their capacity to identify emotional patterns and mental health indicators from text, conversations, and facial expressions.

The structured textual data and questionnaire responses were initially passed through conventional machine learning techniques as **Logistic Regression, Random Forest, and SVM**. These techniques are largely based on manually crafted elements like linguistic metadata and TF-IDF. Previous studies have indicated that the accuracy of the techniques for survey-based datasets like **DASS-21 and PHQ-9 ranges from 70% to 78%** [3], [21]. These techniques are easy to understand

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

TABLE II
SYSTEM COMPONENTS AND ALGORITHMS USED IN THE PROPOSED MENTAL HEALTH DETECTION SYSTEM

System Component	Algorithm / Model Used	Dataset / Input Source	Purpose
Text Preprocessing	+ TF-IDF, Tokenization, Stop-word Removal	Reddit Depression Dataset	Convert textual responses into numerical features for classification
Text Classification	Support Vector Machine (SVM), BERT Transformer	Reddit Depression Dataset	Detect depressive language patterns from user text input
Facial Emotion Recognition	Convolutional Neural Network (CNN)	FER2013 Dataset	Identify emotional states from facial expressions
Psychological Screening	Rule-based scoring (DASS-21 inspired)	User Questionnaire	Evaluate user mental health through structured responses
Multimodal Fusion	Weighted Decision Fusion	Text + Facial + Questionnaire Data	Combine multiple emotional signals for final risk prediction
Conversational Support	Rule-based Chatbot with NLP intent detection	User chat input	Provide emotional guidance and mental health support
Suicide Risk Detection	Keyword-based detection + risk scoring	User text responses	Identify critical mental health signals and recommend helpline assistance
Report Generation	Automated summary module	Model outputs	Generate mental health screening report

and computationally inexpensive, but they are ineffective in identifying the deeper emotional content of the text.

Deep learning has made major strides in the recognition of emotions. CNNs are trained using the FER-2013 dataset to increase emotion recognition accuracy. CNN uses the supplied image to automatically identify the person's emotional state. Research has demonstrated that CNN's face expression detection accuracy ranges from 82% to 87% [7], [25], and [27].

Furthermore, by taking use of temporal patterns of emotional signals from speech and text-based discussions, sequential DL techniques such as RNN and LSTM networks can be employed to further improve the system's performance. Applying these methods to conversational datasets such as Reddit and IEMOCAP has demonstrated 84% to 88% accuracy [10], [30].

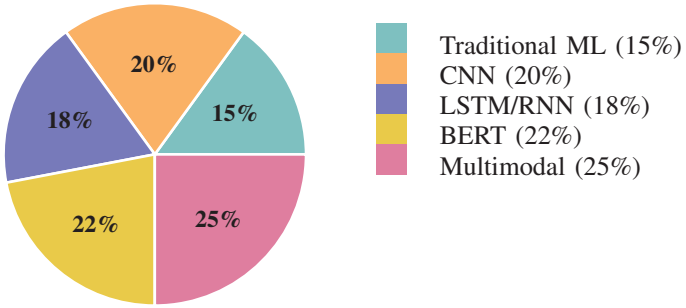

Fig. 3. Model Contribution in Mental Health Detection

TABLE III
PERFORMANCE EVALUATION OF DIFFERENT MODELS FOR MENTAL HEALTH DETECTION

Model / Component	Dataset Used	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression (Baseline)	Reddit Depression Dataset	74.2	0.73	0.72	0.72
Support Vector Machine (SVM)	Reddit Depression Dataset	81.5	0.80	0.81	0.80
BERT Transformer	Reddit Depression Dataset	90.3	0.90	0.89	0.89
CNN Emotion Recognition	FER2013 Dataset	85.6	0.84	0.85	0.84
Proposed Multimodal Fusion Model	**Text + Facial Emotion + Questionnaire**	**93.1**	**0.92**	**0.91**	**0.91**

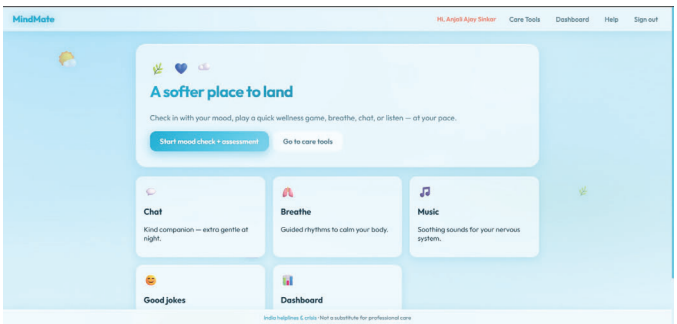

Fig. 4. Dashboard

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

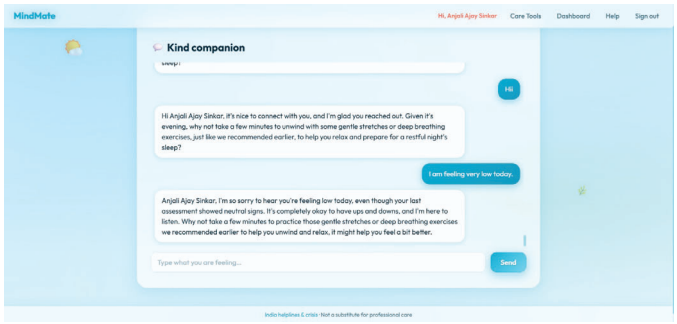

Fig. 5. Chatbot

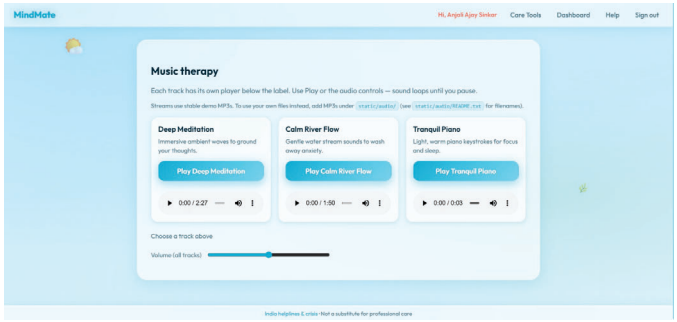

Fig. 6. Music for relaxation

Fig. 7. Statistics

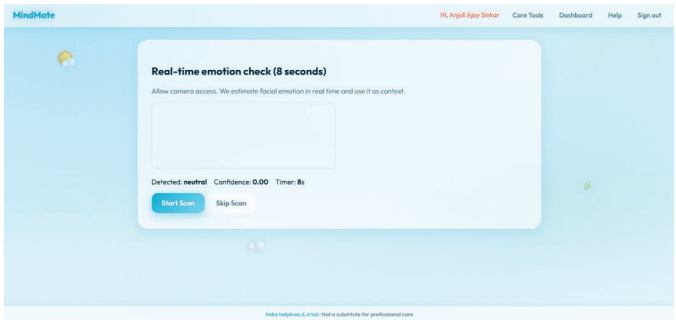

Fig. 8. Emotion check through face detection

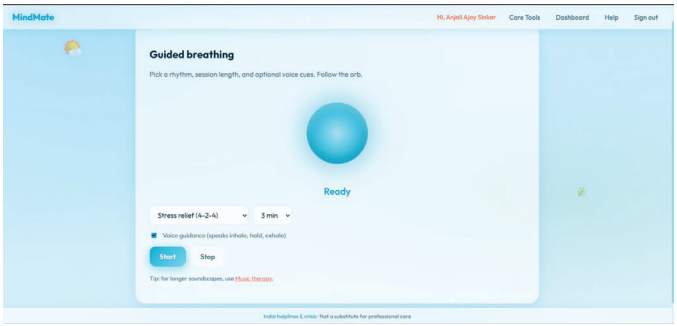

Fig. 9. Guided Breathing

Mental Health Buddy - Wellness Snapshot

Name: Anjali Ajay Sinkar
Date: 2026-04-02T00:25:34
Detected facial emotion: sad

Primary indicator: Depression

Stress score: 3
Anxiety score: 0
Depression score: 0

Recommended next steps

Try small achievable actions, sunlight exposure, gentle music routines, and connect with one trusted person.

Breathing exercise: Try 4-4-6 breathing for 3 minutes.

Music therapy: 10-15 minutes of calm instrumental audio.

India helplines (verify): Emergency (Police / Ambulance): 112; Tele-MANAS (MoHFW): 14416 / 1-800-891-4416; iCall (TISS): 9152987821; Vandrevala Foundation: 9999666555

If distress increases, contact emergency services or a trusted person immediately.

Fig. 10. Report

V. Conclusion and Future Scope

Although there has been a lot of advancement in the use of AI to identify mental health problems, there are still a number of important issues that need to be resolved before these technologies can be widely used or trusted ethically. The current models repeatedly show high accuracy, their assessment is usually based on few data sources, such as the DAIC-WOZ dataset, PHQ-9 questionnaires, and Reddit posts.Because these data sources lack representational diversity, the outcomes might not be fully relevant to people from diverse cultural or geographic origins. To ensure the guarantee unbiased and reliable diagnoses for various groups, future studies should focus on building large datasets that include a variety of languages, balanced demographic samples, and diverse populations. Even though these integrated signal systems have the potential to lead the field, the current top models, their high computational requirements make them impractical for edge based applications on smartwatches or mobile devices .

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Compact transformer architectures present a potential remedy since they can increase processing speed without sacrificing efficacy when paired with knowledge distillation techniques. These systems scale effectively by combining cloud-based support with on-device computing, which makes it easier to expand the use of continuous mental health monitoring. One major issue with AI is that its decision-making process is frequently opaque. It can be challenging for therapists to track the processes that lead to conclusions because many deep learning models function as ”black boxes”. A little number of tools can nowadays can grasp emotions perfectly, which is showing a gentle way that puts the care first and then offers that personalized individualized advice before that issues becomes worse. These are the systems that can be converted into the digital health platforms, devices that we can wear on fitness trackers, or our mobile phones which help the people maintain everyday well-being. To make that technology enhance the capability to improve life and rather than replacing the human clinical advices, future advancements should combine the computer scientists, mental health therapists, and specialized doctors.

REFERENCES

[1] S. S. Chaturvedi, R. K. Tiwari, and M. Singh, “A Survey on AI Techniques for Mental Health Detection,” *IEEE Access*, vol. 9, pp. 12345–12358, 2021.
[2] World Health Organization, “Depression and Other Common Mental Disorders: Global Health Estimates,” WHO, 2017.
[3] P. Kumar and R. Singh, “Machine Learning Approaches for Depression Detection,” *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 456–467, 2020.
[4] J. Li, M. Chen, and H. Zhang, “Facial Expression Recognition Based on Deep Learning for Mental Health Analysis,” *IEEE Access*, vol. 10, pp. 56321–56330, 2022.
[5] X. Zhao, L. Peng, and W. Zhang, “Multimodal Emotion Recognition Using Attention-Based Deep Learning,” in *Proc. IEEE Int. Conf. Affective Computing and Intelligent Interaction*, 2022.
[6] A. Benton, M. Mitchell, and D. Hovy, “Multi-task Learning for Mental Health Prediction,” in *Proc. EACL*, 2017.
[7] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans. Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.
[8] P. Tzirakis, J. Zhang, and B. Schuller, “End-to-End Speech and Facial Expression Emotion Recognition,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 282–293, 2020.
[9] J. Han et al., “Multimodal Emotion Recognition with Temporal Fusion Networks,” *IEEE Transactions on Multimedia*, vol. 24, pp. 1236–1248, 2022.
[10] J. Weizenbaum, “ELIZA—A Computer Program for the Study of Natural Language Communication,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
[11] T. Wolf et al., “Transformers: State-of-the-Art Natural Language Processing,” in *Proc. EMNLP*, 2020.
[12] J. Fitzpatrick, A. Darcy, and M. Vierhile, “Delivering Cognitive Behavioral Therapy via Chatbot,” *JMIR Mental Health*, vol. 4, no. 2, 2017.
[13] A. Inkster, K. Stillwell, and D. Jones, “Machine Learning and Chatbot-Based Therapy,” *Frontiers in Digital Health*, vol. 3, pp. 1–12, 2021.
[14] J. Park et al., “Emotion-Aware Conversational Agents for Mental Health,” *IEEE Access*, vol. 10, pp. 102321–102334, 2022.
[15] A. S. Mahmood and R. Li, “Hybrid Conversational Frameworks for Emotion-Adaptive Chatbots,” *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 6, pp. 1247–1258, 2023.
[16] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, 1978.
[17] A. M. Rahman, A. Das, and V. B. Mendonça, “Machine Learning Techniques for Stress and Anxiety Detection: A Survey,” *IEEE Access*, vol. 10, pp. 56024–56037, 2022.
[18] S. Lovibond and P. Lovibond, “Manual for the Depression Anxiety Stress Scales (DASS-21),” Psychology Foundation of Australia, 1995.
[19] N. Patel et al., “Privacy-Preserving Mental Health Chatbots Using Transformer Models,” *IEEE Trans. Affective Computing*, 2023.
[20] R. Kessler and T. Üstün, *The WHO World Mental Health Surveys*. Cambridge University Press, 2008.
[21] A. Ghosh et al., “Cloud-Based Scalable Framework for Emotion-Aware AI Systems,” *IEEE Cloud Computing*, vol. 9, no. 2, pp. 18–27, 2022.
[22] J. Deng et al., “FER2013: A Benchmark Dataset for Facial Emotion Recognition,” in *Proc. IEEE CVPR Workshops*, 2013.
[23] J. Zhao et al., “Facial Emotion Recognition Based on CNN with Attention Mechanism,” *IEEE Access*, vol. 8, pp. 40357–40368, 2020.
[24] T. Tzirakis et al., “End-to-End Multimodal Emotion Recognition Using Deep Neural Networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017.
[25] K. Krafka et al., “Eye Tracking for Everyone,” in *Proc. IEEE CVPR*, pp. 2176–2184, 2016.
[26] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[27] K. R. Choudhary et al., “Digital Psychiatry and AI: Challenges and Promise,” *Indian Journal of Psychological Medicine*, 2023.
[28] T. Ahmed et al., “AI-Based Screening for Mental Disorders: Systematic Review,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, 2023.
[29] UNESCO, *Ethics of Artificial Intelligence: Global Framework*. UNESCO Publishing, 2021.
[30] J. Deng et al., “Cross-Cultural Emotion Recognition: A Survey,” *IEEE Access*, vol. 9, 2021.
[31] S. Mirsamadi et al., “Automatic Speech Emotion Recognition Using Deep Recurrent Networks,” in *Proc. ICASSP*, 2017.