iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84259

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

AI for Mental Health: Audio Sentiment Analysis

R. Rajesh[1], Sankula Sireesha[2]
[1]*Assistant Professor* , [2]*MCA Final Semester, Master of Computer Applications, Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India*

Abstract: Depression is becoming a major mental health issue, and many people don't realize they have it because the symptoms can be hard to notice. The way a person talks can actually reveal a lot about how they're feeling emotionally and mentally. This system is designed to detect depression by analysing both how someone speaks (like their tone, pitch, and rhythm) and what they say (their sentiment or emotions in words). The process starts with speech-to-text (STT) technology, which turns spoken words into written text.

Then, Natural Language Processing (NLP) methods examine the text to understand its emotional tone. At the same time, advanced audio processing looks at how the speech sounds by extracting features like Mel-frequency cepstral coefficients (MFCCs), which help capture the characteristics of the voice. These different pieces of information both the way someone speaks and the meaning of their words are combined and analyzed using an advanced deep learning model called LSTM. This model has been trained on a well-balanced dataset to ensure accurate results. By studying patterns in both speech and text over time, the system can effectively detect signs of depression.

I. INTRODUCTION

Emotions are a mix of mental and physical reactions that come from changes in the brain and body. They shape our thoughts, feelings, and actions, helping us react to different situations in life. Emotions can be of various types, such as anger, fear, happiness, sadness, and even depression.

These feelings influence how people experience life and interact with others. Depression, in particular, is more than just feeling sad for a short time. It is a serious mental health condition that brings constant feelings of sadness, a lack of interest in daily activities, and a general drop in well-being. Simple tasks may start feeling exhausting, and staying focused or motivated becomes difficult. Depression can be triggered by different factors, such as genetics, stressful events, or medical conditions. Some common signs include feeling hopeless, being constantly tired, struggling to focus, changes in appetite and sleep, and avoiding social interactions. Detecting depression early is important because it allows people to get the help they need before things get worse. Identifying symptoms early can improve mental health and prevent long-term struggles. Thanks to technology and medical advancements, depression can now be identified in different ways, including analyzing speech, facial expressions, and patterns in writing. Spreading awareness and encouraging people to seek professional help can make a big difference in improving their overall quality of life.

II. EXISTING SYSTEM

Most depression detection systems focus on how a person talks instead of what they say. They listen to things like voice tone, pitch, and rhythm to check for signs of depression.

But they don't pay attention to the actual words people use or the emotions behind them. This means they might miss important clues, like when someone often uses negative words, speaks without excitement, or sounds dull. Because of this, these systems might not be as accurate, since what someone says is just as important as how they say it. Another limitation of these systems is that they mainly depend on audio features, which may not always provide a complete understanding of a person's mental state. A person's voice can change due to factors such as illness, stress, or background noise, which may affect the accuracy of the results. Most existing models also struggle to perform well on imbalanced datasets where the number of depressed and non-depressed samples is unequal. In addition, many systems do not combine different types of information, such as speech and text, to improve prediction accuracy. As a result, the overall performance and reliability of depression detection remain limited, creating the need for a more comprehensive and accurate approach.

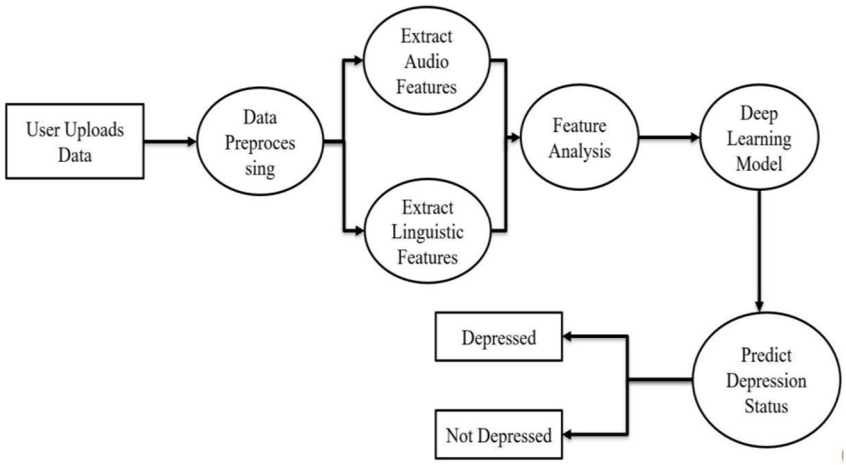


Figure : Flow Chat

A. Challenges

Based on the Final Documentation (Depression Detection Using Audio and Linguistic Features), here are the main Challenges in the Existing System. These are derived from the limitations discussed in the literature survey and system analysis.

1) Lack of Linguistic Analysis: Most existing depression detection systems focus only on audio features such as tone, pitch, and rhythm while ignoring the actual words spoken by the user. Since emotions and mental state can also be reflected through language, relying only on speech characteristics may reduce detection accuracy. This limits the system's ability to understand the complete emotional context.

2) Dependence on a Single Data Source: Many existing approaches depend only on voice recordings for detecting depression. They do not combine multiple sources of information, such as speech patterns and textual sentiment, to make predictions. As a result, important indicators of depression may be missed, leading to less reliable outcomes.

3) Imbalanced Dataset Problem: Depression datasets often contain significantly more non-depressed samples than depressed samples. This imbalance makes it difficult for machine learning models to learn equally from both classes. Consequently, the model may become biased toward the majority class and fail to accurately identify depressed individuals.

4) Difficulty in Real-Time Detection: Several existing models have been developed and tested only in research environments and are not optimized for real-time applications. Processing audio, extracting features, and generating predictions can take considerable time, making them less suitable for continuous monitoring or instant depression screening.

5) Sensitivity to Noise and Speech Variations: The performance of many depression detection systems is affected by background noise, poor audio quality, different accents, speaking speeds, and pronunciation variations. These factors make it difficult to extract meaningful speech features consistently, reducing the overall accuracy and reliability of the system in real-world scenarios.

III. PROPOSED SYSTEM

The current system checks for depression by analyzing how a person speaks, focusing on their tone, pitch, and rhythm. However, it doesn’t pay attention to the actual words they use. This is a drawback because what people say can also reveal important signs of their mental state. The new system improves this by combining both how a person speaks (audio features) and what they say (linguistic features like sentiment). It uses the DAIC-WOZ dataset and LSTM technology to make detection more accurate.

IV. ALGORITHM

The proposed system uses Deep Learning techniques to detect depression by analyzing both audio and linguistic features. The system processes speech recordings by extracting important audio characteristics such as MFCC features and combines them with linguistic information obtained through speech-to-text and sentiment analysis. An LSTM (Long Short-Term Memory) model is then trained using these combined features to classify whether a person is depressed or not. By integrating multiple sources of information, the system improves prediction accuracy and provides more reliable depression detection.

A. Feature Extraction

The system begins by extracting meaningful information from the input speech. Audio features such as pitch, tone, rhythm, and MFCC coefficients are extracted using audio processing techniques. At the same time, the speech is converted into text, and sentiment analysis is performed to identify the emotional content. These extracted features serve as the input for the classification model.

B. LSTM Model

The extracted audio and linguistic features are provided as input to an LSTM network. LSTM is capable of learning long-term relationships in sequential data, making it suitable for analyzing speech patterns. It processes the combined features and predicts whether the user is likely to be depressed or not.

C. Training Process

The model is trained using labeled data obtained from the DAIC-WOZ dataset. During training, the network continuously updates its parameters using the Binary Cross-Entropy loss function and the Adam optimizer to minimize prediction errors. This iterative learning process improves the overall performance and accuracy of the model.

D. Supervised Learning

The proposed system follows a supervised learning approach where labeled audio samples are used to train the LSTM model. The dataset is divided into training and validation sets, allowing the model to learn speech patterns associated with depression and evaluate its performance before testing.

E. Handling Class Imbalance

Since depression datasets usually contain fewer depressed samples than non-depressed samples, the system applies the SMOTE technique to balance the dataset. This helps the model learn equally from both classes and reduces prediction bias during training.

F. Model Evaluation

After training, the model is evaluated using testing data. Performance metrics such as Accuracy, F1-Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are calculated to measure the effectiveness of depression detection and ensure reliable predictions.

V. TECHNIQUES

The proposed system combines several Artificial Intelligence, Deep Learning, and Natural Language Processing techniques to detect depression accurately. These techniques work together to process speech, extract useful information, classify emotional patterns, and generate reliable predictions. By integrating multiple techniques, the system improves both accuracy and robustness in real-world depression detection.

A. Speech-to-Text (STT)

Speech-to-Text technology converts spoken audio into text using the Whisper model. This allows the system to analyze the spoken content in addition to voice characteristics, providing valuable linguistic information for depression detection.

B. Natural Language Processing (NLP)

Natural Language Processing is used to analyze the transcribed text and determine the sentiment expressed in the user's speech. The extracted sentiment helps identify emotional patterns that may indicate depression and complements the audio-based analysis.

C. Audio Feature Extraction

The system extracts Mel-Frequency Cepstral Coefficients (MFCCs) and other speech characteristics such as tone, pitch, and rhythm using the Librosa library. These audio features capture important voice patterns that are commonly associated with depression.

D. LSTM Deep Learning Model

The extracted audio and linguistic features are combined and processed using an LSTM-based Deep Learning model. LSTM effectively learns sequential speech patterns and produces the final prediction indicating whether the individual is depressed or not.

VI. METHODOLOGY

The methodology of the proposed system follows a structured workflow for detecting depression using both audio and linguistic information. Initially, the user provides a speech recording as input. The audio is preprocessed to remove noise and improve quality before extracting speech features such as MFCCs. Simultaneously, the speech is converted into text using Speech-to-Text technology, and sentiment analysis is performed to identify emotional characteristics.

The extracted audio and linguistic features are combined into a unified feature set. Before training, the dataset is balanced using SMOTE to reduce class imbalance. The processed features are then supplied to an LSTM model, which learns temporal speech patterns and predicts whether the person is depressed or not. Finally, the trained model is evaluated using metrics such as Accuracy, F1-Score, MAE, and RMSE to ensure reliable and accurate depression detection.

A. Outputs

```
floating-point round-off errors from different computation orders. To turn them off, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0
2025-02-19 10:16:41.437706: I tensorflow/core/util/port.cc:153] oneDNN custom operations are on. You may see slightly different numerical
2025-02-19 10:17:49.634566: I tensorflow/core/platform/cpu_feature_guard.cc:210] This TensorFlow binary is optimized to use available CPU
To enable the following instructions: AVX2 AVX512F AVX512_VNNI FMA, in other operations, rebuild TensorFlow with the appropriate compiler
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you tra
✅ Model loaded successfully!
Model Input Shape: (None, 1, 14)
✅ Class Weights: {0: 0.7121212121212122, 1: 1.6785714285714286}
2/2 ━━━━━━━━━━━━━━━━━━━━ 1s 364ms/step

✅ Best Threshold: 0.4500
✅ Test Accuracy: 0.5957
✅ Test F1 Score: 0.5366

🔍 Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.52      0.64        33
           1       0.41      0.79      0.54        14

    accuracy                           0.60        47
   macro avg       0.63      0.65      0.59        47
weighted avg       0.72      0.60      0.61        47

✅ Test MAE: 0.4772
✅ Test RMSE: 0.4798
>>>
```

Figure 1: Testing Result

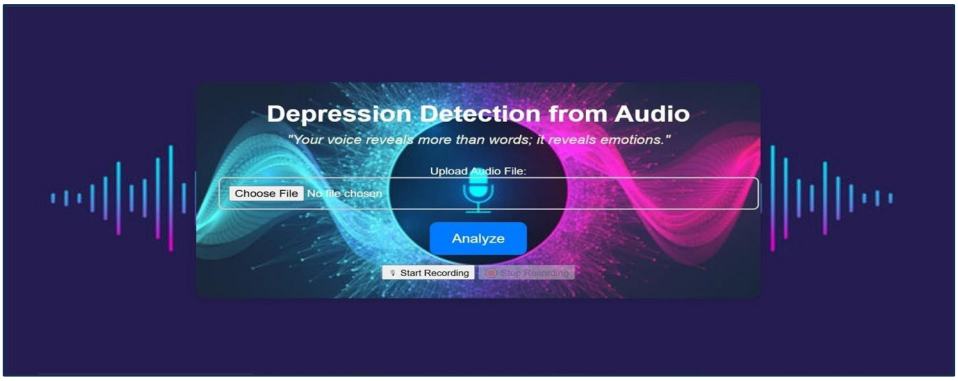


Figure 2 : Input page

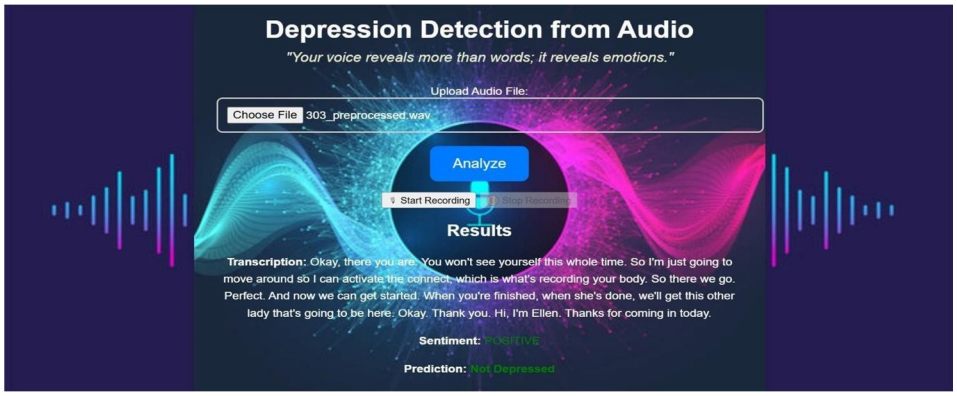


Figure 3 :Output Page-1

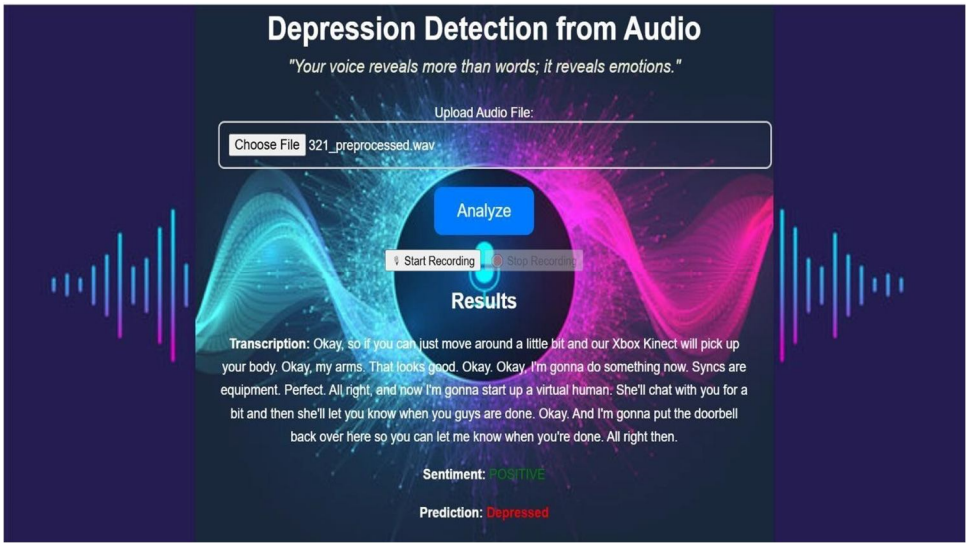


Figure 4 : Output Page-2

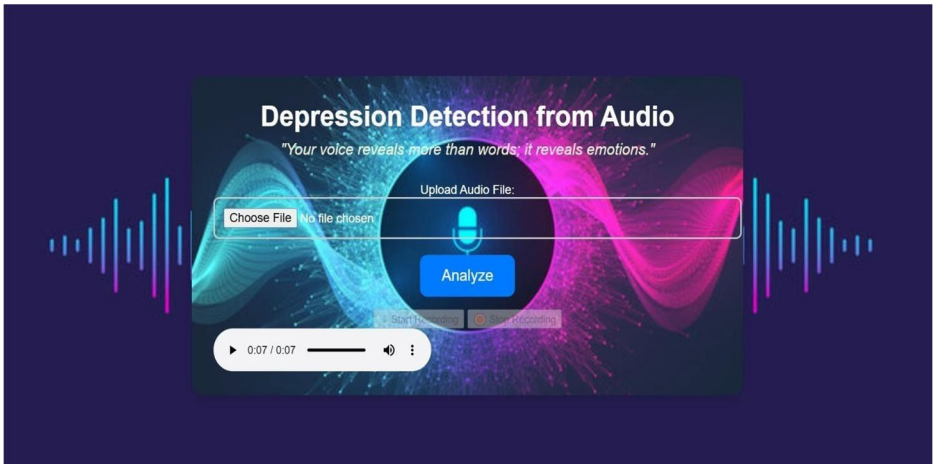


Figure 5 : Output Page-3

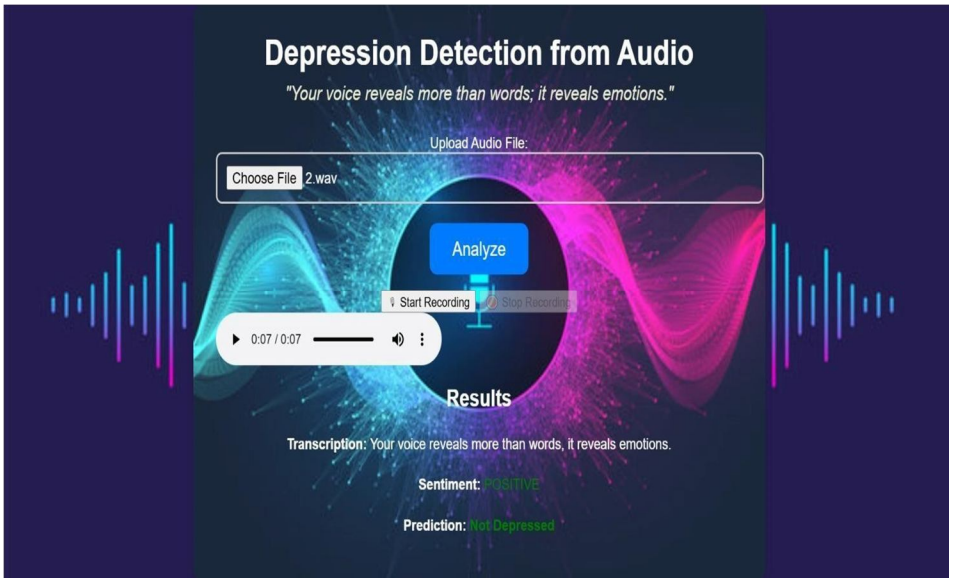


Figure 6: Output Page-4

VII. CONCLUSION

IdeaCrafter successfully addresses the challenges faced by aspiring entrepreneurs by providing an intelligent and scalable mentorship platform. The integration of AI, machine learning, and modern web technologies enables real-time, personalized guidance. The system improves accessibility, reduces costs, and enhances decision-making for users. It also provides a strong foundation for future enhancements such as mobile applications, voice-based interaction, and advanced AI integration.
Overall, IdeaCrafter contributes to building a more inclusive and innovative entrepreneurial ecosystem.

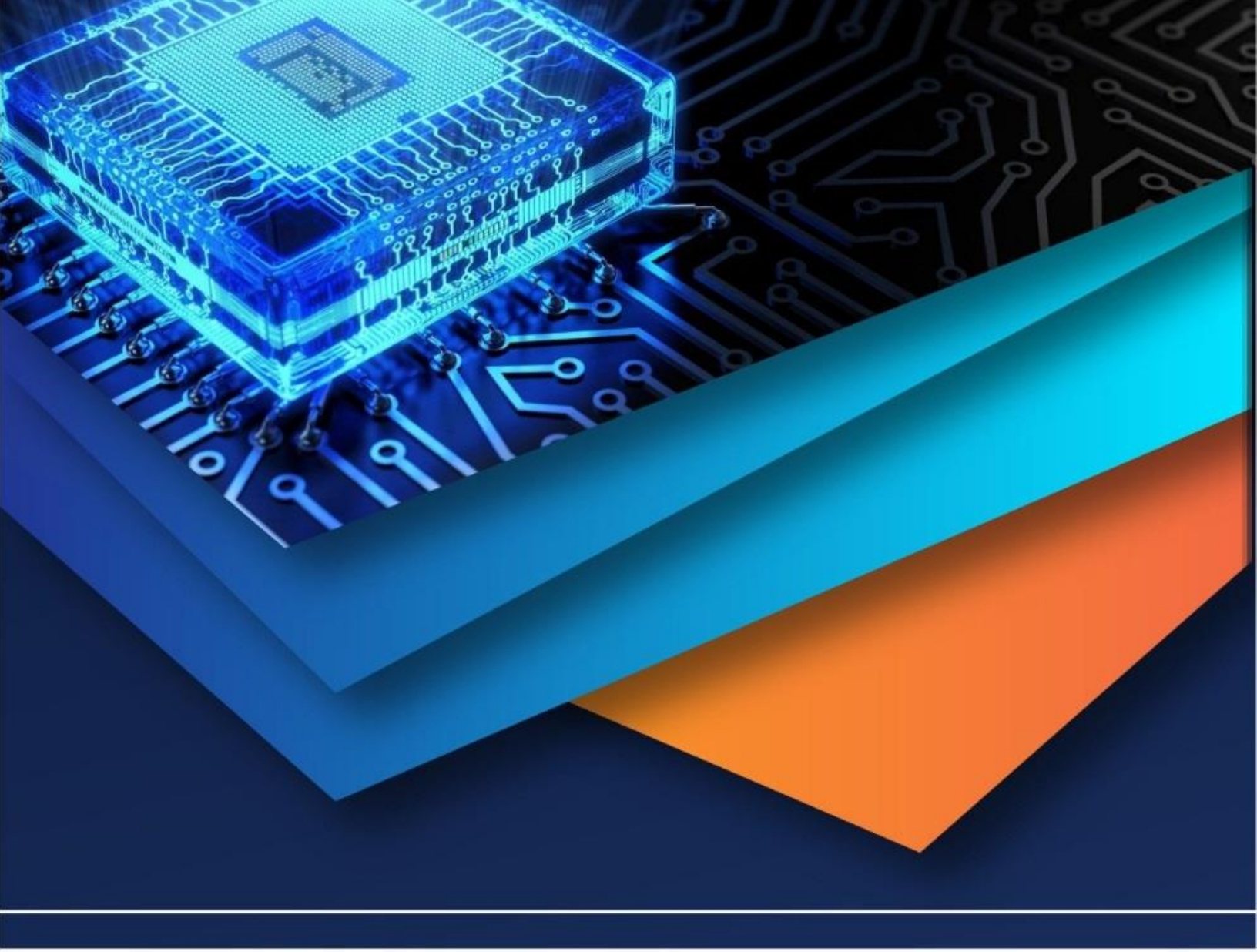
REFERENCES

[1] A. K. Das and R. Naskar, "A Deep Learning Model for Depression Detection Based on MFCC and CNN Generated Spectrogram Features," Biomedical Signal Processing and Control, vol. 90, p. 105898, 2024.
[2] A. Othmani, et al., "Towards Robust Deep Neural Networks for Affect and Depression Recognition from Speech," in Pattern Recognition. ICPR International Workshops and Challenges, Springer International Publishing, 2021.
[3] A. Saidi, S. B. Othman, and S. B. Saoud, "Hybrid CNN-SVM Classifier for Efficient Depression Detection System," in Proceedings of the 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET), IEEE, 2020.
[4] H. Negi, et al., "A Novel Approach for Depression Detection Using Audio Sentiment Analysis," International Journal of Information Systems & Management Science, vol. 1, no. 1, 2018.
[5] I. Madhavi, et al., "A Deep Learning Approach for Work-Related Stress Detection from Audio Streams in Cyber-Physical Environments," in Proceedings of the 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, 2020.

BIBLOGRAPHY

R.Rajesh working as an Assistant Professor in Computer Science and Engineering at Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh. With 3 years of experience in Computer Science and Engineering (B.tech), accredited by NAAC, he has a membership in IAENG with his area of interest in Java Full Stack, DBMS, Data Structures, and C Language.

Sankula Sireesha is currently pursuing his final semester of Master of Computer Applications (MCA) at Sanketika Vidya Parishad Engineering College, which is accredited with an 'A' grade by NAAC, affiliated to Andhra University, and approved by AICTE. With a keen interest in Artificial Intellegence , has undertaken her postgraduate project titled " AI For Mental Health: Audio Sentiment Analysis." The project is carried out under the guidance of R. Rajesh Assistant Professor , SVPE.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY