Development of a Depression Detection System Using Speech and Text Data

B Surekha Reddy, Jishitha Kondaveti, V Akshaya Bhavani and P Aishwarya

May 18, 2023

Development of a Depression Detection System using Speech and Text Data

[1]Dr. B. Surekha Reddy
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
b.surekhareddy@iare.ac.in

[2]Kondaveti Jishitha
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
dimpydkj@gmail.com

[3]V Akshaya Bhavani
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
akshayavaradhi@gmail.com

[4] P Aishwarya
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
19951A0405@iare.ac.in

Abstract*— Depression is a mental health disorder that affects a significant portion of the global population. However, early detection and intervention are crucial for effective treatment. Current depression detection models often rely on a single modality, either audio or text data, which may lead to inaccurate results due to the limited information being considered. In this paper, we propose a novel approach for the accurate detection of depression in individuals using both audio and text data. This approach addresses the limitations of existing solutions by combining objective measures of speech and text data, providing a non-intrusive and efficient means for depression detection. In this project, two models were trained: a speech emotion recognition model based on Long Short-Term Memory (LSTM) and a text model based on the Random Forest Method. The TESS dataset was used to train the speech emotion recognition model, whereas the text model was trained on the Twitter dataset, which is accessible on Kaggle. The results indicate that our proposed approach is highly effective in detecting depression, with an accuracy of 98% achieved by the speech emotion recognition model and 95% by the depression detection model on the testing data. The outputs of both models are combined using a decision tree method, resulting in an accuracy of 100%. The proposed method employs a Decision Tree algorithm that takes the output of both the Speech Emotion Recognition (SER) model and the text-based depression detection model as inputs and applies a set of rules to classify users into one of three categories: depressed, mildly depressed, or not depressed. Then a webpage was developed where users can input their speech and text data and receive a prediction of their depression status based on the integrated output of both models. Overall, the results demonstrate that the proposed system provides a promising solution for the early detection of depression.

Keywords*—Depression, Mental Health, Early Detection, LSTM, Random Forest, Non-intrusive, Decision Tree*

I. Introduction

Depression is an alarming medical disorder that affects millions of people worldwide. According to the WHO [1], about 300 million people worldwide suffer from depression. According to another study [2], it predicts that by the year 2030, major depressive disorder will become the second highest contributor to disability on a global scale. Although it is not easily identified, it has significant and severe effects. Clinical expertise is currently employed in addition to questionnaire surveys to identify depression. Early depression diagnosis and evaluation are difficult, and their accuracy is heavily dependent on patient cooperation and medical professional skill. People nowadays use social platforms such as Reddit, Pinterest, Twitter, and Instagram to express their emotions, thoughts, and provide details about their daily life . Due to its widespread usage as a mode of communication, textual data exhibits several characteristics that render it highly suitable for both sentiment analysis, chatbots, natural language generation and many more**.**

Speech emotion recognition can indeed contribute to understanding depression by analyzing the emotional cues conveyed through speech patterns and intonations. By accurately identifying specific emotional states, such as sadness, apathy, or despair, this technology offers insights into a person's mental well-being. Leveraging speech emotion recognition can aid in early detection and intervention, potentially leading to improved support and treatment for individuals struggling with depression. As a result, extracting his/her emotional signal from the speaker is critical for the research of depression. Deep learning algorithms have been used to detect depression using speech and text data, which can improve detection accuracy and efficiency.

II. Literature Survey

Depression is only one example of a mental health issue that is incredibly difficult to handle. There are professional medical treatments accessible, but they take time. The problem with current depression detection models is that they often rely on a single modality, either speech or text data, which can lead to inaccurate results due to the limited information being considered. Additionally, some existing models lack interpretability, making it difficult to understand how the model arrived at its conclusions.

Depression is a major concern, especially after the COVID-19 pandemic. According to [3], there has been a significant increase in mental health disorders such as generalized anxiety, social anxiety, depression, and panic/somatic symptoms, especially among young adults and adolescents. The study emphasizes the need of mental health care and interventions, as well as the use of technology and telemedicine, in providing critical therapies during and after the epidemic. As a result, it is critical to undertake research and develop models for the detection and management of depression in order to improve people's overall mental health and well-being.

Recent research [4] has shown that the use of audio and text data can be an effective method for detecting depression. Audio data, such as speech and ambient sounds, can provide information about an individual's emotional state and can be used to detect changes in their speech patterns, such as a monotone voice, slower speech rate, and reduced pitch range. Textual information, such as social network posts and text messages, has the ability to provide vital insight into a person's thoughts and emotions. It can also be used to detect changes in their writing style and linguistic patterns.

Natural language processing techniques have recently gained favor in diagnosing depression from social media data due to the extensive use of online communication and the prospect of identifying persons who may be at risk for depression. The study [5] employed a word list to identify depression tendencies in individual tweets and two common classification models to predict depression class, Multinomial NB and SVM. The Multinomial NB performed well with an accuracy of 83% and an F1-score of 83.29%, while the SVM performed well with an accuracy of 79% and an F1-score of 79.73%. The study did, however, acknowledge the limitations of supervised learning classification in predicting depression from text data.

Another study [6] used twitter data from 55 Indonesian users, they analyzed it using a grading system and compared to clinical screening labels from psychologists. Results showed comparable accuracy and F1 score values between regular and text-specific processing, with the latter slightly outperforming. However, the small sample size is a drawback of this study, highlighting the necessity for large datasets in efficient machine learning techniques. In recent years, ML and DL techniques have been increasingly used to detect depression. The study [7] used ML algorithms to predict anxiety, depression, and stress levels based on the DASS 21 questionnaire. Five different algorithms were utilized, with the RF classifier having the highest accuracy scores ranging from 71.4% to 79.8%. The NB algorithm had the highest accuracy across all three scales.

Depression is a complex and multifaceted disorder that can manifest in different ways, including through text, speech, and other modalities. Therefore, a multi-modal approach that combines data from various sources, such as text, audio, and physiological signals, could offer a more thorough and accurate method of detecting depression. Several studies have shown that multi-modal depression detection algorithms have great potential. For example, in [8], researchers created a model that used data from speech and facial expression analysis to diagnose depression with high accuracy.

The study [9] adopted a multi-modal technique to detect depression that incorporated physiological signs, speech, and text data and achieved better results than models that used only one modality. While hybrid models have shown potential in detecting depression from text, more study is needed to enhance detection speed. Furthermore, the use of multi-modal techniques that incorporate data from multiple sources may provide a more thorough and accurate method of detecting depression.

Deep learning algorithms for SER have received a lot of interest recently. The work [10] presents a summary of the several DL approaches utilized for SER, as well as their contributions and limitations. While these methods provide simple model training and the efficiency of shared weights, they also have drawbacks such as extensive layer-wise internal architecture and over-learning during layer-wise information memorization. The study by [11] focuses on recognizing negative emotions in human speech data in Thai language, which is one application of deep learning in SER. The researchers assessed the accuracy of deep learning classifiers in recognizing negative emotions using four open emotional speech datasets. The CONV1D recognized negative emotions with a high accuracy of 96.60%.

Earlier studies in SER largely employed typical machine learning algorithms to classify emotions from speech inputs, such as SVM. Recent research [12] has concentrated on deep learning and neural network-based techniques to increase SER performance. LSTM, GAN, and VAN are the most often utilized deep learning architectures for SER. LSTM, in particular [13], has demonstrated substantial potential in boosting SER accuracy by modelling long-term relationships in speech signals. The way that the memory of LSTM can capture temporal information and manage input sequences of varying durations makes it well-suited for modelling speech sounds. Despite these advances, SER still faces obstacles and constraints, such as a lack of well-designed datasets and uncertainty in emotional annotations.

The study [14] aims to create a new method for detecting depression early on by combining audio and text data. They used the SESE on 160 Chinese people to see how their emotions changed. To diagnose depression, the researchers retrieved low-level audio data and used a multi-modal fusion technique based on Deep Spectrum and word vector features. The proposed model was accurate to 0.912 and had an F1 score of 0.906. The study's shortcomings, however, include its small sample size and lack of variety, implying the need for more detailed research on a larger and more diverse dataset.

In conclusion, multiple studies have demonstrated that using speech and text data to identify depression can be an effective strategy. The addition of both speech and text data can provide a complete knowledge of an individual's emotional state and increase depression diagnosis accuracy. However, more research is required to validate these findings and discover an ideal technique for utilizing speech and text data in the diagnosis of depression.

III. Proposed Method

Depression detection using speech and text data typically involves a multi-modal approach, where both speech and text data are used together to make predictions. The general methodology for detecting depression using speech and text data has the following steps: Data Collection, Data Preprocessing, Feature Extraction, Model Training, Model Evaluation and Model Deployment.

A. Proposed Depression Detection System using Speech Data

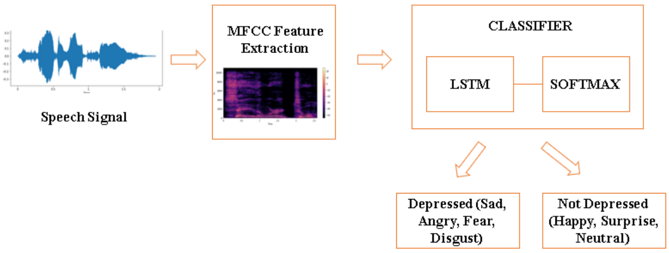

Speech Signal
MFCC Feature Extraction
CLASSIFIER
LSTM
SOFTMAX
Depressed (Sad, Angry, Fear, Disgust)
Not Depressed (Happy, Surprise, Neutral)

Figure 1 Proposed Speech Emotion Recognition Model

The speech emotion recognition (SER) system incorporates essential blocks such as speech signal processing, data collection, feature extraction, and classification. Speech signal processing involves techniques like noise removal and frequency enhancement using pre-processing and Librosa. Data collection relies on the availability of large datasets, and in this case, the TESS dataset with 2800 audio files is utilized. Feature extraction involves extracting relevant features from speech signals, with Mel Frequency Cepstral Coefficients (MFCC) being a popular technique. MFCC reduces variability and represents the audio signal's properties using the Mel frequency scale and Discrete Cosine Transform (DCT). Finally, the classification block employs LSTM networks to assign emotional labels to input speech signals based on the extracted features.

- **LSTM Architecture**

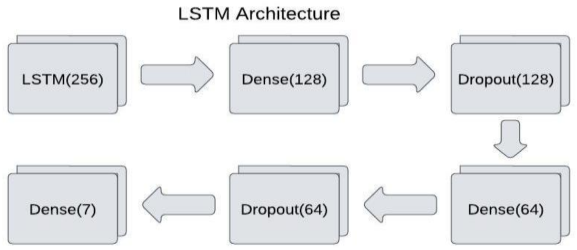


Figure 2 LSTM Layers for Speech Emotion Recognition Implementation

LSTM model is sequential model and built using keras in python. LSTM stands for long short-term memory networks, used in the field of DL. It belongs to the family of RNNs and is specifically designed to capture long-term dependencies, particularly in tasks involving sequence prediction. The structure of an LSTM model consists of multiple LSTM cells connected in a sequential manner to handle input sequences. Each LSTM cell incorporates three gates: input, forget, and output gates, which regulate the information flow within the cell. These gates enable the LSTM network to effectively process and understand sequential data by learning and remembering long-term dependencies between different time steps.

B. Proposed Depression Detection System using Text Data

Depression detection text models use NLP techniques to analyze text samples and identify signs of depression. The training phase of a depression detection text model involves feeding the model a large dataset of labeled examples, where the labels indicate whether a given text sample is indicative of depression or not. This enables the model to learn patterns and characteristics linked with depression. The testing phase of a depression detection text model involves using the trained model to predict whether new, unseen text samples are indicative of depression or not.

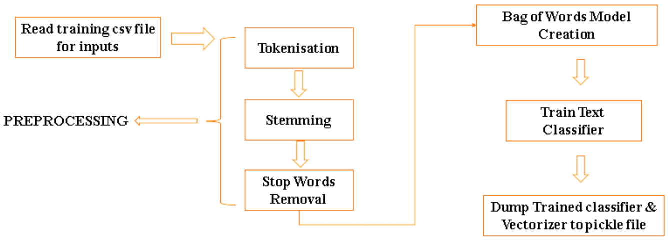

Read training csv file for inputs
PREPROCESSING
Tokenisation
Stemming
Stop Words Removal
Bag of Words Model Creation
Train Text Classifier
Dump Trained classifier & Vectorizer to pickle file

Figure 3 Training Phase of Text Model using Random Forest Classifier

In the training phase of a text classification model using a Random Forest classifier, pre-processing is performed to transform the raw text data into a suitable format for analysis. This includes text cleaning, tokenization, Stopwords removal, and lemmatization or stemming. These steps prepare the text data for the classifier and improve its accuracy. A bag of words model is then created, which represents the text data in a numerical format for analysis. The Random Forest classifier uses this model to categorize new text input based on learned features. The classifier is trained by adjusting hyperparameters and evaluated using various metrics. If the performance is unsatisfactory, hyperparameters can be fine-tuned or a different algorithm can be tried. Finally, the trained classifier and vectorizer are dumped into a pickle file for easy preservation and later deployment.

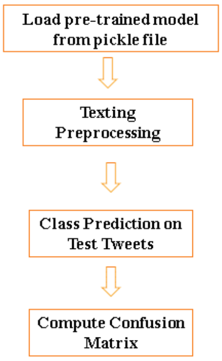

Load pre-trained model from pickle file
Texting Preprocessing
Class Prediction on Test Tweets
Compute Confusion Matrix

Figure 4 Testing Phase of Text Model using Random Forest Classifier

In the testing phase of a text model, a pre-trained model is loaded from a pickle file using Python's pickle module. The loaded model is then used to preprocess the random input text by applying techniques such as tokenization, stemming, and stop-word removal to structure the text for analysis. The preprocessed text is fed into the model to generate predictions on whether the text indicates depression or not. Class prediction on test tweets involves categorizing unseen tweets based on their content using the pre-trained machine learning model. The preprocessed test data is vectorized using TF-IDF Vectorizer, and the model uses it to forecast the classes of the test tweets. The resulting predictions are compared to the true labels of the test data to compute a confusion matrix, which consists of true positives, true negatives, false positives, and false negatives. The confusion matrix serves as the foundation for evaluating the model's performance using metrics like accuracy, precision, recall, and F1 score, providing insights into its effectiveness in detecting instances of depression.

- **Random Forest Classifier**

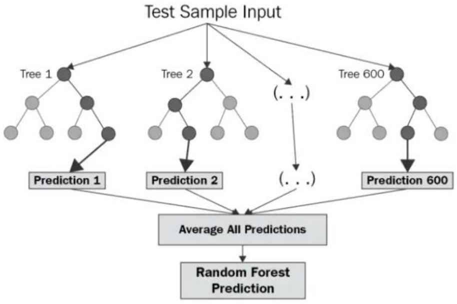


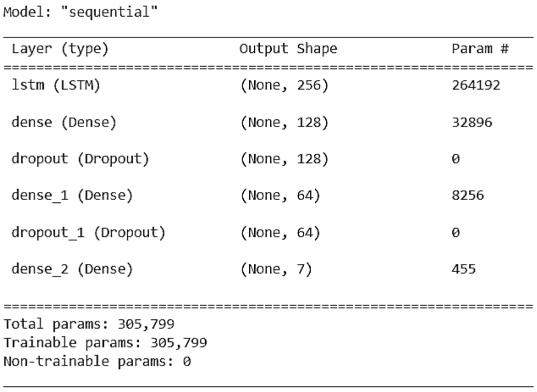
Figure 5 Random Forest Architecture

RF is a powerful ML technique that is commonly used in text categorization tasks like detecting depression in tweets. The dataset is initially preprocessed in this method, and then features are retrieved using techniques such as Bag-of-Words, TF-IDF, and POS. The RF classifier is then trained using the preprocessed dataset. The technique involves creating multiple decision trees using different subsets of the training dataset in the training phase. These decision trees collectively contribute their votes to determine the final class of test objects.

IV. Implementation

A. *Implementation of Depression Detection Model with Speech Data using LSTM*

The proposed method for the speech model utilizes the Toronto Emotional Speech Dataset (TESS), consisting of 2800 audio files with spoken words representing different emotions(anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). The model employs an LSTM architecture, capable of learning and retaining information from long input time series. The architecture includes layers such as Dense, Dropout, ReLU, and Softmax, which contribute to classification and nonlinearity. After building and compiling the LSTM model, it is fitted using a training dataset to adjust the weights and optimize its performance. The proposed LSTM-based speech model follows an 80:20 train-test data split. This split ensures that the model is trained on a substantial amount of data while also having unseen data for evaluation. Additionally, the model is trained over 50 epochs, indicating that the training process iterates 50 times over the training data.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 256)               264192

 dense (Dense)               (None, 128)               32896

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 64)                8256

 dropout_1 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 7)                 455

=================================================================
Total params: 305,799
Trainable params: 305,799
Non-trainable params: 0
_________________________________________________________________
```

Figure 6 Sequential Model of LSTM Model

The model is then evaluated using metrics like accuracy, f1 score, precision, and recall assessing its ability to detect emotions from speech. Predictions can be made on new speech signals using the trained model, and performance assessment is crucial in determining its accuracy. The model's performance is evaluated on a test set of speech recordings, comparing predicted sentiments to real emotions.

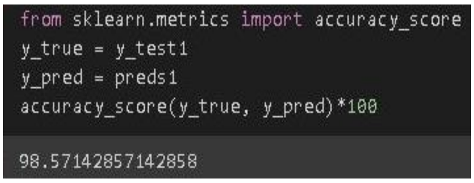
B. *Implementation of Depression Detection Model with Text Data using Random Forest*

The implementation of the text model for depression detection involved several stages. Firstly, a dataset of over 10,000 tweets labeled as depressed or non-depressed was obtained from Kaggle. The dataset was specifically curated for text-based depression detection models and covered a wide range of tweet samples. The text data underwent preprocessing, which included removing punctuation, stop words, and converting text to lowercase. Tokenization was performed to transform the tweets into numerical vectors for input features. Feature extraction techniques, such as the bag-of-words model and TF-IDF, were used to represent the tweets as feature vectors, capturing the importance of words. Parts of Speech (POS) tagging was also incorporated to enhance the model's understanding of grammatical patterns. The Random Forest Classifier was trained and evaluated using a 75:25 train-test split. Evaluation metrics such as accuracy, precision, recall, F1 score, and the confusion matrix were employed to assess the model's performance.

V. Results And Discusssion

A. *Emotion Recognition using Speech Data*

The accuracy of the SER model, which was trained using an LSTM model is displayed in Figure 7.



```
from sklearn.metrics import accuracy_score
y_true = y_test1
y_pred = preds1
accuracy_score(y_true, y_pred)*100

98.57142857142858
```

Figure 7 Accuracy for Speech Emotion Recognition using LSTM Model

Figure 8 shows the confusion matrix obtained for speech emotion recognition model that uses LSTM classifier.

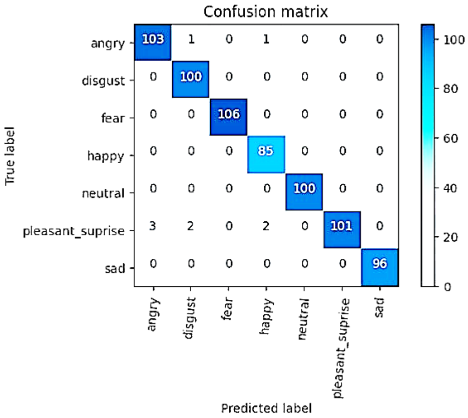


Figure 8 Confusion Matrix for Speech Emotion Recognition using LSTM Model

Table 1 summarizes the evaluation metrics for classifying seven emotions. The main findings indicate consistently high precision scores of 1.00 for all emotions, indicating minimal false positives. The majority of emotions achieved a recall score and F1 score of 1.00, demonstrating accurate identification of true positives.

Table 1 Precision, Recall, F1 Score for Depression Detection using Speech Data

Emotion	Label	Precision	Recall	F1 Score
Sad	0	1.00	0.98	0.99
Disgust	1	1.00	0.99	0.99
Fear	2	1.00	1.00	1.00
Happy	3	1.00	1.00	1.00
Neutral	4	1.00	1.00	1.00
Surprise	5	0.98	1.00	0.99
Angry	6	0.99	1.00	0.99

Figure 9 displays the graph for the accuracy of the training and validation data during the model training process.

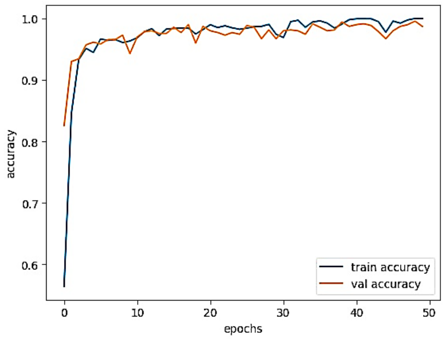
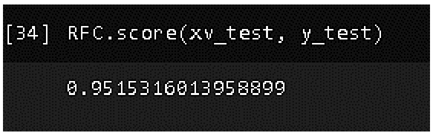


Figure 9 Training and Validation Accuracy Graph for Speech Emotion Recognition Model

B. *Depression Detection Model using Text Data*

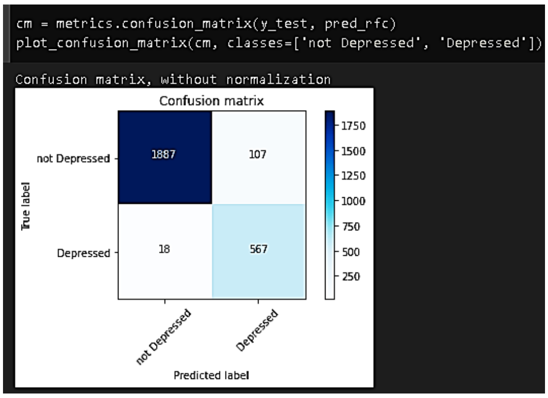
The Text model is trained using Random Forest Model, the accuracy obtained is shown in Figure 10 and also the confusion matrix can be seen in Figure 11 respectively. Figure 12 shows the validation and training accuracy graph for depression detection using text data.



```
[34] RFC.score(xv_test, y_test)

     0.9515316013958899
```

Figure 10 Accuracy for Depression Detection Model using Text Data



```
cm = metrics.confusion_matrix(y_test, pred_rfc)
plot_confusion_matrix(cm, classes=['not Depressed', 'Depressed'])

Confusion matrix, without normalization
```

Figure 11 Confusion Matrix for Depression Detection Model using Text Data

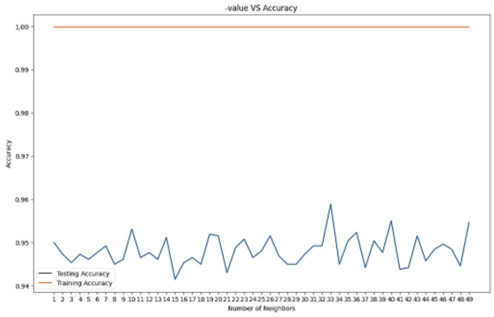


Figure 12 Training and Validation Accuracy Graph for Depression Detection Model using Text Data

Table 2 presents evaluation metrics for a Random Forest model used to detect depression through text data. The model exhibits a precision of 0.99, indicating high accuracy in identifying true positives, while its f1 score of 0.96 signifies a balanced performance in correctly classifying depressed individuals.

Table 2 Precision, Recall, F1 Score and Accuracy for Depression Detection Model using Text Data

Algorithm	Precision	Recall	F1 score	Accuracy
Random Forest	0.99	0.94	0.96	0.95

Table 3 provides a comprehensive overview of the performance and methodologies employed in previous studies for SER, alongside the proposed method. Table 4 offers a comparative analysis between previous studies and the proposed method for text modeling.

Table 3 Comparison of Proposed Method with Existing Methods for Speech Emotion Recognition (SER)

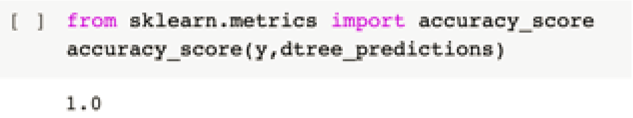
	Emotions Detected	Database	Method	Accuracy (%)
Guizzo et al. [27]	Anger, Happy, Sad, Neutral, Disgust, Pleasant Surprise, Fear	TESS	CNN	53.05
Vege et al. [28]	Angry, Sad, Happy, Neutral		RNN	75
Zahra et al. [29]	Happy, Anger, Sad, Pleasant Surprise Neutral, Disgust, Fear		DNN	89.96
Proposed Method with Speech Data	Happy, Fear, Sad, Pleasant Surprise Neutral, Disgust, Anger		**LSTM**	**98**

Table 4 Comparison of Proposed Method with Existing Methods for Depression Detection Model using Text

	Method	Dataset	Accuracy (%)
Kumar et al. [30]	Multinomial Naïve Bayes	Twitter	77.89
	Gradient Boosting		79.12
	Ensemble Vote Classifier		85.04
Ahmad et al. [31]	Naïve Bayes		71
	KNN		72
	SVM		79
Proposed Method with Text Data	**Random Forest**		**95**

C. Depression Integration of Speech Text Model using Decision Tree Model

The speech emotion recognition model and depression detection text model were integrated into a decision tree model to determine a user's depression status. The decision tree model achieved a remarkable accuracy of 100%, indicating its effectiveness in accurately predicting depression. In Figure 13, it shows the accuracy of decision tree model.



```
[ ]  from sklearn.metrics import accuracy_score
     accuracy_score(y,dtree_predictions)

     1.0
```

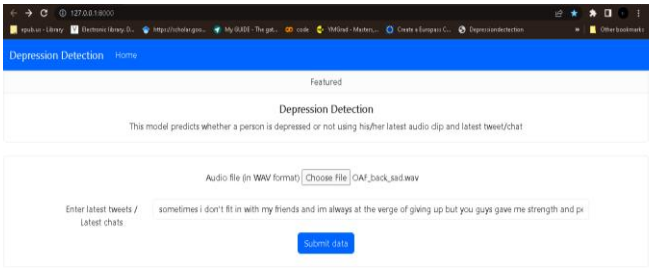
Figure 13 Accuracy Score for Speech- Text Integrated Model using Decision Tree Algorithm

Table 5 illustrates the collective findings of the depression detection system that combines speech and text inputs.

Table 5 Combined Results of Speech-Text Depression Detection System

Input		Output
Speech	**Text**	
Sad, Fear, Angry	Depressed Text	Depressed
Neutral, Happy, Surprise	Non- Depressed Text	Not- Depressed
Sad, Fear, Angry	Non- Depressed Text	Mildly Depressed
Neutral, Happy, Surprise	Depressed Text	Mildly Depressed
Other	Other	Mildly Depressed

The output is presented in a clear and concise format, classifying the user into one of three categories: depressed, mildly depressed, or not depressed. Figure 14 shows the website interface of depression detection and Figure 15 shows the output page of the website.



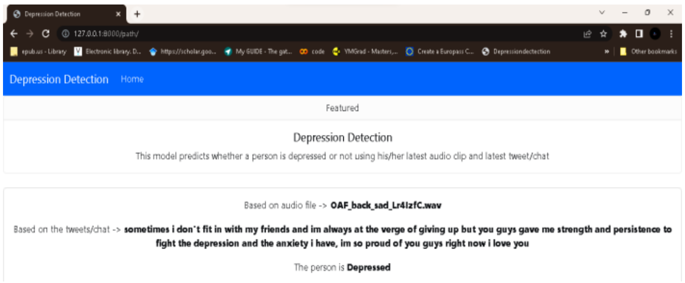
Depression Detection

This model predicts whether a person is depressed or not using his/her latest audio clip and latest tweet/chat

Audio file (in WAV format) Choose File OAF_back_sad.wav

Enter latest tweets / Latest chats: sometimes i don't fit in with my friends and im always at the verge of giving up but you guys gave me strength and pe

Figure 14 Depression Detection Website Interface



Depression Detection

This model predicts whether a person is depressed or not using his/her latest audio clip and latest tweet/chat

Based on audio file -> **OAF_back_sad_Lr4IzfC.wav**

Based on the tweets/chat -> **sometimes i don't fit in with my friends and im always at the verge of giving up but you guys gave me strength and persistence to fight the depression and the anxiety i have, im so proud of you guys right now i love you**

The person is **Depressed**

Figure 15 Website Output Display Page

VI. CONCLUSION

In this project, we developed a model that accurately predicts an individual's mental state based on their speech and text inputs. The speech emotion recognition model achieved 98% accuracy, while the text classification model achieved 95% accuracy in identifying depression. By integrating these models and using a decision tree, we achieved a remarkable accuracy of 100% in classifying individuals as depressed, mildly depressed, or not depressed. The user-friendly website further enhances accessibility, making it a valuable tool for those seeking support with depression.

VII. FUTURE WORK

In future work, there are several areas of improvement for the depression detection model. Firstly, incorporating additional features such as facial expressions and gait, alongside speech, can enhance the accuracy of the model by capturing more comprehensive information about an individual's mental state. Secondly, exploring different neural network architectures and feature engineering techniques can further optimize the current model, improving its performance and prediction capabilities.

ACKNOWLEDGMENT

We would like to thank Institute of Aeronautical Engineering (IARE) college for conducting this research project and Dr. B Surekha Reddy for supervising us throughout the project.

REFERENCES

[1] Depression Key Facts [Last accessed on 2021 Nov 9]. Available from: http://www.who.int/news-room/fact-sheets/detail/depression.

[2] Izutsu, Takashi, AtsuroTsutsumi, Harry Minas, Graham Thornicroft, Vikram Patel, and Akiko Ito. "Mental health and wellbeing in the Sustainable Development Goals." The Lancet Psychiatry 2, no. 12 (2015): 1052-1054.

[3] Hawes, Mariah T., Aline K. Szenczy, Daniel N. Klein, Greg Hajcak, and Brady D. Nelson. "Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic." Psychological medicine 52, no. 14 (2022): 3222-3230.

[4] Solieman, Hanadi, and Evgenii A. Pustozerov. "The detection of depression using multimodal models based on text and voice quality features." In 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), pp. 1843-1848. IEEE, 2021.

[5] Deshpande, Mandar, and Vignesh Rao. "Depression detection using emotion artificial intelligence." In 2017 international conference on intelligent sustainable systems (iciss), pp. 858-862. IEEE, 2017.

[6] Ji, Shaoxiong, Shirui Pan, Xue Li, Erik Cambria, Guodong Long, and Zi Huang. "Suicidal ideation detection: A review of machine learning methods and applications." IEEE Transactions on Computational Social Systems 8, no. 1 (2020): 214-226.

[7] Shah, Faisal Muhammad, Farzad Ahmed, Sajib Kumar Saha Joy, Sifat Ahmed, Samir Sadek, RimonShil, and Md Hasanul Kabir. "Early depression detection from social network using deep learning techniques." In 2020 IEEE Region 10 Symposium (TENSYMP), pp. 823-826. IEEE, 2020.

[8] Anastasia, Olympia Simantiraki, C-M. Vazakopoulou, Charikleia Chatzaki, Matthew Pediaditis, Anna Maridaki, Kostas Marias et al. "Facial geometry and speech analysis for depression detection." In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1433-1436. IEEE, 2017.

[9] Orabi, Ahmed Husseini, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. "Deep learning for depression detection of twitter users." In Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic, pp. 88-97. 2018.

[10] Abbaschian, Babak Joze, Daniel Sierra-Sosa, and Adel Elmaghraby. "Deep learning techniques for speech emotion recognition, from databases to models." Sensors 21, no. 4 (2021): 1249.

[11] Shewalkar, Apeksha, Deepika Nyavanandi, and Simone A. Ludwig. "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU." Journal of Artificial Intelligence and Soft Computing Research 9, no. 4 (2019): 235-245.

[12] Ye, Jiayu, Yanhong Yu, Qingxiang Wang, Wentao Li, Hu Liang, Yunshao Zheng, and Gang Fu. "Multi-modal depression detection based on emotional audio and evaluation text." Journal of Affective Disorders 295 (2021): 904-913.

[13] Lipton, Zachary C., John Berkowitz, and Charles Elkan. "A critical review of recurrent neural networks for sequence learning." arXiv preprint arXiv:1506.00019 (2015).

[14] Guizzo, Eric, Tillman Weyde, and Jack Barnett Leveson. "Multi-time-scale convolution for emotion recognition from speech audio signals." In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6489-6493. IEEE, 2020.