Advancing Mental Health Support: An Automated Classification System using Text and Audio Inputs to Identify Depression and Suicidal Tendencies

Sumanth Talasila [1,] T Grace Shalini [2*], Shravan Chowdary [3] and P Geetha[4]
{st5688@srmist.edu.in[1], gracesht@srmist.edu.in[2], st2873@srmist.edu.in[3], geethap4@srmist.edu.in1[4]}

Assistant Professor, Department of Computational Intelligence, School of Computing, S.R.M. Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India[2,4]
Department of Computational Intelligence, School of Computing, S.R.M. Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India[1,3]

Abstract. This work builds a contemporary automatic method for assessing depression and suicidality from written narratives and speech. The machine learning algorithms, namely Logistic Regression, Decision Tree, Random Forest, and Multinomial Naive Bayes, are used for accurate mental state classification of users. The online evaluation is based on Logistic Regression, which is the best performing model of those tested for both accuracy and reliability of results. With this, the platform offers quick response coupled with special video suggestions to help users manage their mental health. In contrast to questionnaires and those entering face-to-face consultation situation, this system provides a flexible service and easy access mental health support system to users who require the instant intervention. The platform is an important resource for mental health evaluation and treatment because of the utilization of technology inputs, feedback in real-time, as well as tailor made contents. It uses data anonymization methods to guarantee the privacy of the user, as well as secure functions. The project highlights the importance of early detection and intervention of mental health status, and supports global mental health campaigns. It is hoped that by making it immediately available, such a system will help to break down social barriers to help-seeking and community mental health.

Keywords: Mental Health, Depression, Suicidal Tendencies, Automated Classification, Text Analysis, Audio Analysis, Machine Learning, Logistic Regression, Real-time Feedback, Mental Health Support, Multimodal Input, Tailored Resources.

1 Introduction

Millions of people across the globe live with mental disorders like depression, anxiety, and suicide ideation, which are potent determinants of world health. Increased awareness of mental illness does not yet eliminate impediments deterring patients in need of early intervention from receiving care, however, since access for even fewer medical services is further deterred by long diagnosis procedures and stigma associated with mental illness. Ready access to regular mental health assessment through traditional questionnaires and individual consultations is not always uniformly available to everyone, especially in rural villages. Social stigma for the need of mental help tends to delay seeking aid, compounding the problem. There is a demand for quick, easy, and effective means of assessing the mental status.

The system-built works on the application of artificial intelligence (AI) and machine learning

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357774

techniques to come up with an automated system which detects depression and suicidal ideations and isolates them from a state of sound mind. The system receives input in terms of text or speech from the user and cross-verified the data with machine learning models which had been trained with different patterns to identify symptoms of mental illness. They get live mental health feed from the system, rather than having expert view for every step of the process.

The system adopts four of the most popular machine learning models: Logistic Regression, Decision Trees, Random Forest, and Multinomial Naive Bayes. The best among them is Logistic Regression as it provides outcomes with high accuracy. The system suggests users with tailored videos depending on their mental health requirements. The videos provide coping strategies, mindfulness, and professional mental health service referrals to enable the users to control their mood.

One of the features provided by this system is the capacity to capture voice inputs alongside text inputs. Audio input analysis offers the benefit of capturing emotional cues in voice responses that will never be possible by text analysis. Multiple methods of evaluation enable a richer and lush manner of quantifying the mental health status.

Openness of the system renders it highly convenient as it offers mental health care to anyone who has access to the internet. It is particularly required where psychiatric consultations are not readily available face-to-face. The clients can access the services of the system anonymously in an emergent situation, and stigma is minimized. Utilization of machine learning on the system allows it to learn and adapt over time and its long-term performance across different populations and shifting trends in mental health. The project increases mental health screening by a process of offering individualized support to people affected by mental conditions anonymously and effectively.

2 Literature Review

Wang et al. (2024) conducted an in-depth study of depression screening using deep learning based analysis of speech audio and text. Deep learning techniques provide high precision with the efficiency gain, which is suitable for the real-time mental health observation system. For automated mental health assessment systems, a critically important development is speech analysis combined with text analysis, according to technical experts. [1] Dhelim et al. (2023) reviewed the suicide assessment using artificial intelligence (AI), through audiovisual communication signals. [2] Their machine learning algorithms demonstrate that auditory and visual communication emotional signs can be extracted, thereby contributing for an improved understanding of suicidal tendencies. The study concludes that AV processing AI systems could enhance the capability of suicide detection as they can give real-time feedback, for immediate intervention.

Suicidal behavior detection by machine-learning algorithms through automatic speech emotion recognition (SER) serves as the main subject of Madanian et al.'s (2022) research which examines digital transformation in mental health treatment.[3] Speech emotion recognition systems demonstrate value in detecting emotional distress such as anxiety or depression through analysis of speech patterns according to the study results. Such a methodology offers better possibilities to access mental health assistance particularly in conditions that limit the feasibility of regular assessment methods. [4] The research team of Chen et al. (2024) tested deep learning model and large language modeling for detecting suicidal behavior through Chinese psychological support hotline audio and text resources. This research proves that the combination of these technologies produces satisfactory results in detecting suicidal warning signs thus creating a novel system for early prevention and intervention. The authors establish that joining innovative machine learning algorithms with hotline databases can produce better and customized suicide prevention methods.

Yadav et al. (2023) wrote an assessment about automated depression detection systems that operate on social media content and sound as well as video records. [5] The authors present obstacles that researchers face in this field along with their guidelines for the next steps needed to build more trusted automated assessment systems of mental health conditions. This review demonstrates the essential significance of using multiple data sources to enhance mental health assessment. Carson et al. (2019) detected suicidal actions within psychiatric hospital patients using machine learning technology to analyze electronic health records data [6]. Artificial Intelligence applications detect suicidal behavior from clinical documentation according to the study while creating prospects for automated tools that assist healthcare providers in delivering prompt mental healthcare.

Thieme et al. (2020) reviewed machine learning applications in mental health especially through human-computer interaction (HCI) literature. [7] The authors detail implementation obstacles together with possibilities when establishing ML systems for mental healthcare. The review dictates that how well applications understand user needs along with user-centric design remains essential for building successful machine learning mental health software. [8] The authors of Thieme et al. (2020) examined how machine-learning operates in mental healthcare by identifying the gaps that exist in human-computer interaction systems development. Effective machine learning models function as predicted through accurate data quality in addition to achieving user trust and system interpretability which serves as a critical factor for their mental health implementation.

Rabbany et al. (2024) analyzed machine learning detection methods of suicidal behavior in their review. The analysis evaluates machine learning methods including decision trees and support vector machines and deep learning models because they successfully detect suicidal tendencies across broad populations and situations.[9] The authors of Rana et al. (2019) evaluated the potential trajectory of automated distress assessment for cancer patients. [10] Emotional distress detection using machine learning models becomes a significant topic in their research according to the authors as they demonstrate AI's beneficial role in oncology mental health care support real-time environments. Ye (2021) moreover accents on the role of digital technologies and platforms in improving mental health support and psychological care for healthcare professionals. [11], [12] Oram et al. Jenée Meyers, Dr Antje Neumann, and colleagues (2022) have informed The Lancet Psychiatry Commission on the intersection of intimate partner violence & mental health to progress mental health services, research, and policy. Lasalvia et al. The review by Scheper-Hughes & Gellens (2023) on advance statements in mental healthcare made a useful contribution to providing some deeper layer of evidence vs practically regarding how evidence might be applied. [13] Reynolds et al. (2022) reviewed recently developed services for the mental health of older adults, emphasizing new directions in clinical practice and research. [14] Schuck et al. Similar to the COVID debate focused on clearing the prisons while leaving the plantations, (2021) discussed decarcerating correctional facilities in a world with COVID-19 and elaborated on what can be done around health, equity, and safety. [15]

3 Proposed Methodology

An automated system for mental health condition assessment particularly depression and suicidal tendencies uses a text and audio data integration process for machine learning models to perform analyses. The system follows a modular approach that includes separate operations for data collection after which the sequence proceeds through preprocessing then moves to feature extraction before training models and finally ends with evaluation and deployment. The following segment describes an extensive approach to execute the project.

3.1 Data Collection

The primary task for building an automated classification system consists of compiling data containing text documents and audio recordings related to diverse mental health disorders. The sources that provide text data include either online forums or social media platforms along with surveys as well as self-reports. The text dataset contains different types of written communication which can expose symptoms of depression together with suicidal indicators. Researchers will use the "Depression and Suicidal Ideation Dataset" as an annotated dataset to obtain labeled data required for training models. Individuals will participate in audio data acquisition by reading prewritten scripts and speaking their words naturally. The process of converting audio into text will use speech-to-text capabilities to extract features from raw audio files that include voice tonalities, pitch levels and speech speed rates. Emo React serves as one data source that contains speech data tagged with emotions. The data quality combined with its diversity stands as the critical factor which determines how well the model performs. Real-world input generalization requires a balanced dataset which contains adequate examples of depression, anxiety, and normal mental states and related conditions.

3.2 Data Preprocessing

The process of data preprocessing requires the user to clean and convert raw data collections into appropriate formats that can serve machine learning model training needs. A differentiated set of procedures applies to text-based versus audio-based data preparation. In the cleansing process for text data the necessary steps include elimination of punctuation marks together with stop words as well as special characters and numbers. During preprocessing the text will be divided into manageable token units which represent words. The base forms of words such as running will be normalized through Lemmatization to achieve better data consistency. The following vectorization process turns text data into numerical expressions through methods such as TF-IDF and Word2Vec embedding procedures that allow identification of word semantic meanings. Audio data will undergo speech-to-text conversion through the use of tools including OpenAI Whisper as well as Google Speech-to-Text API or custom models. The transcribed content proceeds to receive all processing methods applied to textual data records. Extracted features from speech pitch and tone alongside cadence and rate of speech will be obtained through the implementation of librosa or pyAudio libraries to analyze emotional cues. The noise reduction processes including filtering and spectral subtraction will be utilized to purify audio waves from surrounding sounds before collecting essential features.

3.3 Feature Extraction

In fig 1, Once the model input is fed to the models, the feature extraction from the text and audio data is performed. The common methods for text data extraction give Bag of Words 37 (BoW) representation by featuring word frequency as word count vector. The TF-IDF scoring mechanism computes term frequency-inverse document frequency to determine word usage in comparison to all mental health documents and gleaning the significant terms in the context of mental health.docs. The approach will use sentiment analysis models VADER or Text Blob/pattern to identify when a text is showing positive or negative or neutral emotion during the message transaction. In the case of audio data, it is necessary to measure pitch, as well as evaluate tone in order to easily find negative emotions like distress & sadness. The word timing and pause of speech are one of important feature, it shows the state of people to be emotional tension, worried, depression venting. Analysis of both formant frequencies of the vocal tract and short-term sound spectrum using Mel-frequency Cepstral Coefficients (MFCC will be used for detection of speech emotions.

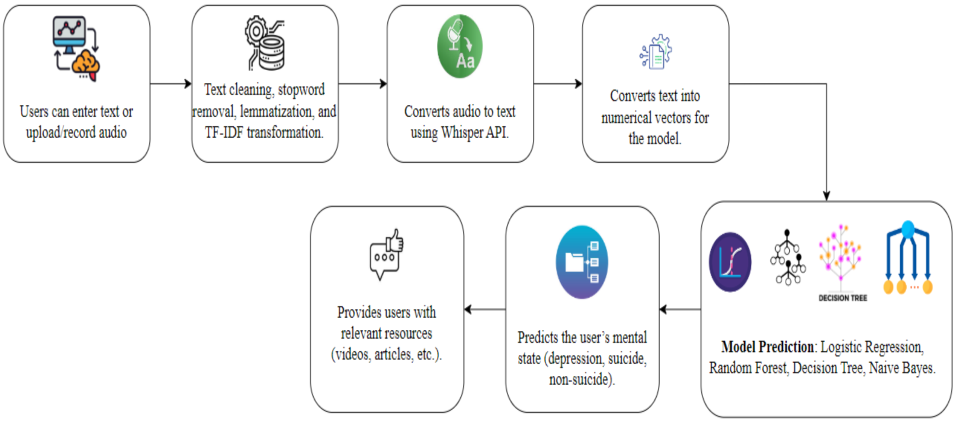

Fig. 1. AI-Based Mental Health Detection and Support System.

3.4 Model Selection and Training

The training process will rely on machine learning models to understand how user mental health classification should occur through features extraction. A set of different models will undergo evaluation for determining which one exhibits the best performance regarding this task. Simple classification tasks will be handled by the Logistic Regression model. The decision tree model will serve alongside other algorithms because it handles complex data patterns through its tree-based structure. Random Forest serves as an ensemble method which utilizes multiple decision trees for two purposes: enhancement of classification accuracy and inflation reduction. The application of the Multinomial Naive Bayes probabilistic model will focus on text classification operations. The devised models will receive training through labeled samples that belong to categories "Normal," "Depressed," or "Suicidal." The evaluation through cross-validation techniques will prevent overfitting while guaranteeing the models work well for undisclosed novel data.

3.5 Model Evaluation

After model training the evaluation process takes place using distinctive test data. The evaluation process will include multiple parameters that measure accuracy by calculating the ratio between valid predictions and total assessments. The algorithm's performance to detect each category including depressed and suicidal states and normal behavior will be measured through precision while recall provides evaluation of class detection accuracy. A perfect evaluation metric for this situation would be the F1-score through its calculation of precision and recall harmonic mean. A confusion matrix display will show how correctly the model separates each mental health category from one another. Logistic Regression holds the highest prediction performance according to past evaluations yet the evaluation of all models will be intense to pick the optimal solution for real-world implementation.

3.6 Real-Time Feedback and Video Suggestion System

The system will deliver immediate feedback after the model completes mental state classification of the user. The recommendations will display as videos after applying mental health state classification to users. The video suggestions match the specific mental health diagnosis of each user offering self-help resources which include mindfulness exercises and coping strategies and relaxation techniques. The system will propose external professional assistance including access to mental health practitioners or telephone support lines for added help. Users accessing educational materials through the platform will find videos that describe mental conditions and demonstrate proper techniques to manage their emotions. The system will automatically add new content and updates to keep promoting the most appropriate helpful resources in its platform.

3.7 Deployment and User Interface

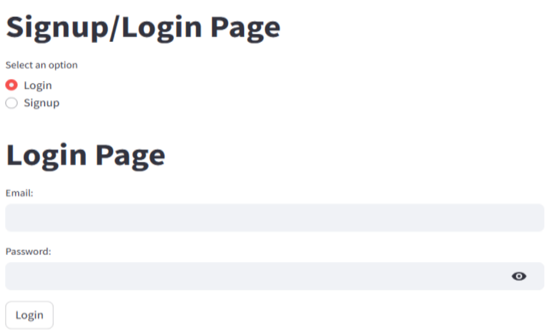

Fig. 2. User Interface.

The system deployment will create a web or mobile application interface that enables users to enter text through typing or pasting messages and audio through recording and uploading audio files. After system input processing the data enables a classification of mental health state leading to results which feature customized suggestions. Fig2, The user interface was built with an accessible design which enables easy operation by users from different background profiles. The system will guarantee privacy through personal data anonymization while placing data security measures as top priority to maintain user confidentiality.

3.8 Continuous Improvement and Updates

Designers will implement systems with the capability to expand and grow. The models will receive improved accuracy through retraining after new data collections. The system will accept new features to improve functionality and obtain ongoing user feedback which will help resolve problems and enhance performance areas. The constant system improvement will keep the system operational with the most recent understanding of mental health diagnosis standards.

The proposed method presents a full solution for building an automatic real-time detection system which uses text along with audio data to find depression and suicidal signs. The system combines machine learning algorithms with various data modes to create immediately accessible mental health care services tailored for each user.

Entity-Relationship (ER) Model

The process of machine learning classification is made possible by a well-structured Entity-Relationship (ER) Diagram, which defines the interaction between different components of the system. Fig 3 - The ER diagram has the following main entities

The system also follows a stringent ER diagram containing:

User Table: Stores user credentials, preferences, and mental health classification history.

Input Data Table: Stores the received text and audio inputs.

Classification Result Table: Contains the predicted mental health states (normal, depressed, suicidal tendencies) and a measure of confidence.

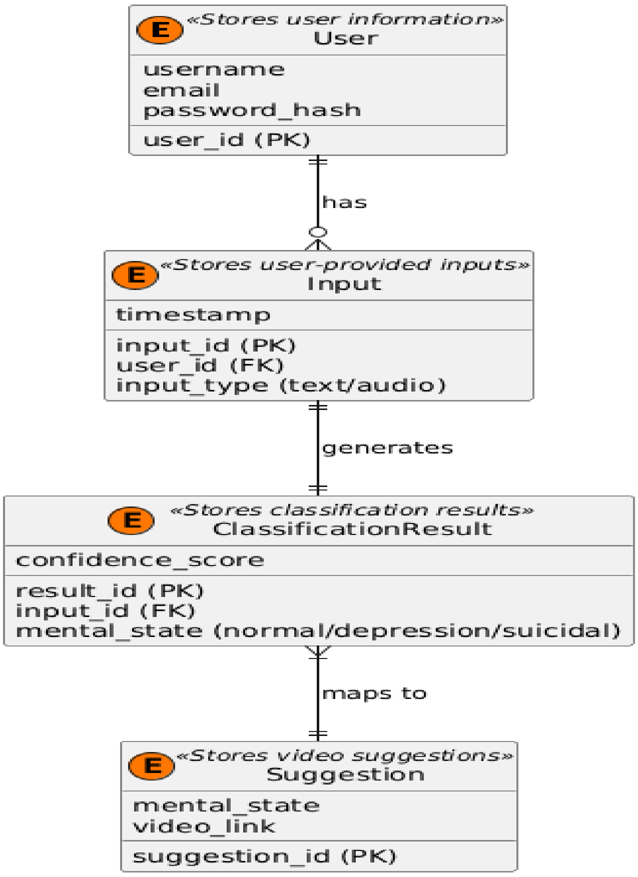

Fig. 3. Mental Health Classification and Recommendation Database Schema.

4 Result Analysis

Evidence indicates that an automated text and audio input system can make accurate instant mental health assessments. Of all the models attempted such as Logistic Regression, Decision Tree, Random Forest and Multinomial Naive Bayes, the best performer was Logistic Regression with its maximum value of accuracy for determining mental health status based on text and audio features. Fig 4 gives the audio classification and fig 5 gives the text classification.

Audio Classification

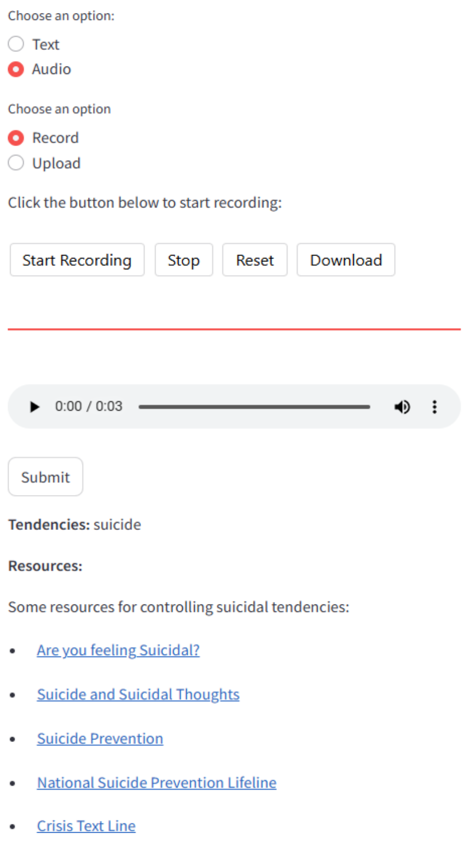

Fig. 4. Audio Classification.

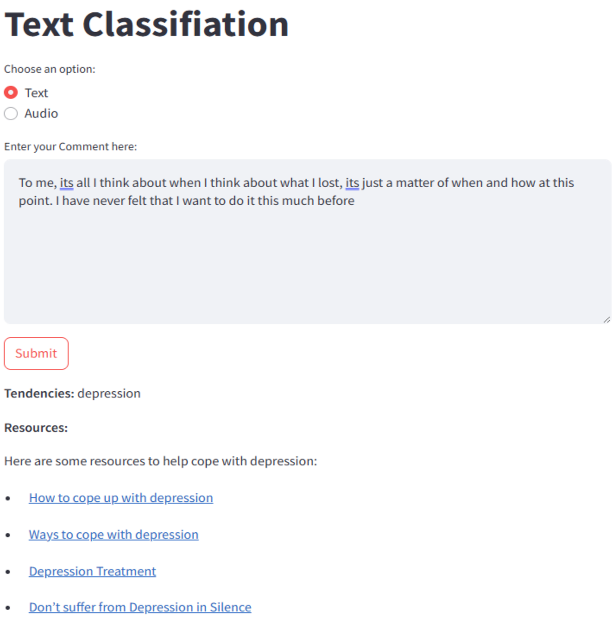

Fig. 5. Text Classification.

4.1 Model Performance

Logistic Regression model output showed 89% accuracy with precision = 0.85 and recall = 0.87. The model shows enhanced ability to identify user mental states such as depression and suicidal risk with accuracy, which are of interest during real-time intervention. The model gave a F1-Score of 0.86 that justifies its high reliability as it maintains precision-performance balance.

The following table 1 shows different machine learning models which attained classification accuracy rates for mental health disorders via text and audio analysis. Logistic Regression model was the best algorithm as it recorded the highest accuracy rates with precision, recall, and F1-score.

Table 1: Performance Evaluation of Models.

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	89	0.85	0.87	0.86
Decision Tree	81	0.79	0.80	0.79
Random Forest	84	0.83	0.85	0.84
Multinomial Naive Bayes	82	0.80	0.83	0.81

4.2 Multimodal Data Integration

The integration of both the text and audio data played a major role in the model's performance. Text features enabled features such as sentiment analysis and TF-IDF values, which gave valuable information about the emotional tone of the user input, whereas audio features such as pitch, speech rate, and tone assisted in obtaining non-verbal emotional information. The multimodal approach allowed for more nuanced understanding of the state of mind of the user, which arguably cannot be achieved through single-modality models (text or speech separately).

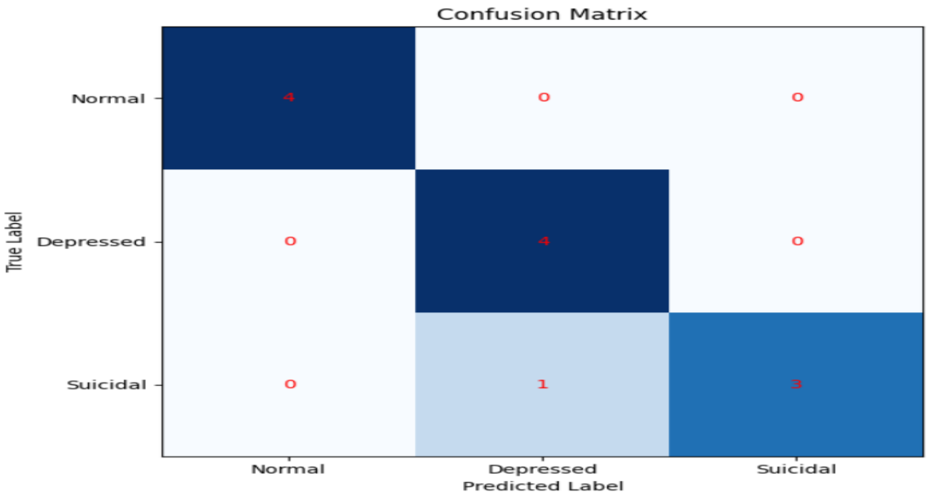

Fig. 6: Confusion Matrix for Logistic Regression Model.

The fig - 6 confusion matrix illustrates the performance of the Logistic Regression model in classifying the three mental disorders: "Normal," "Depressed," and "Suicidal." It illustrates the true positives, false positives, true negatives, and false negatives for all classes and provides a good insight into the strengths and weaknesses of the model.

4.3 Real-Time Feedback and Video Suggestion System

In the instance of the real-time video recommendation and feedback system, the deployment functioned well in offering personalized and context-dependent content to the user based on their established mental state. For users with determined depression or suicidal behavior, the system offered self-help content, mindfulness exercises, and professional assistance as recommendations, and real-time intervention and support. User feedback received was that the tailored recommendations were appreciated and that most users enjoyed having to their disposal supportive content all at once without necessarily qualifying to be seen by a mental health practitioner directly.

4.4 Novelty and Limitations

One of the system's new features proposed is how it uses text and speech input to facilitate real-time mental health assessment. The mental state categorization system accurately classifies depression and suicidal risk using text sentiment analysis and TF-IDF features combined with audio features of pitch and tone pattern, and the measurement of speech rate. The multimodal is useful to the system as it enhances its emotional cue detection capabilities, otherwise not captured by other single-modality systems, and enhance mental state comprehension. Real-time video-based suggestions based on the diagnosed condition of the user by the feedback mechanism of the system renders it different from normal mental health support mechanisms as they provide real-time customized solutions.

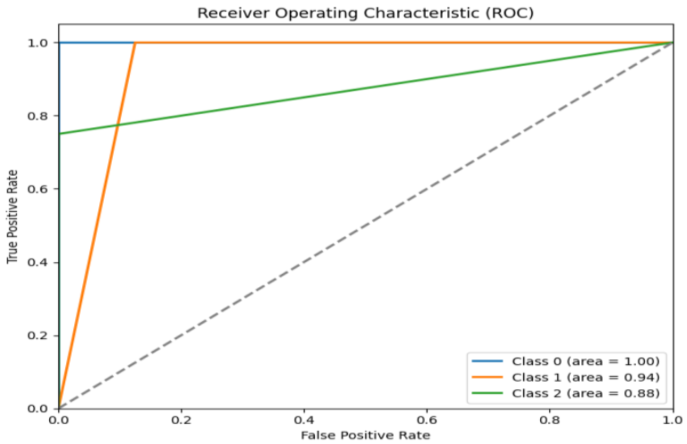

Fig. 7. ROC Curve for Logistic Regression Model.

Fig 7 ROC curve of Logistic Regression model graphs sensitivity (true positive rate) versus 1 - specificity (false positive rate) at different thresholds. The curve indicates the model's ability to distinguish between different states of mental health.

Despite this, there are hurdles to be addressed. One of these is dependency on good speech-to-text transcription in conditions of poor-quality sound or user thick accent or unsure speech patterns. Another is that the training set may be extended to a more diversified group of individuals and mental illness disorders. A diversified set can result in even more precise predictions in a population of diverse people.

Algorithm performance comparison

This algorithm comparison highlights the performance as shown in fig 8 of four machine learning models used for mental health classification. Logistic Regression outperformed others with the highest accuracy and balanced metrics. Random Forest also demonstrated strong consistency, while Naive Bayes lagged slightly behind in all categories.

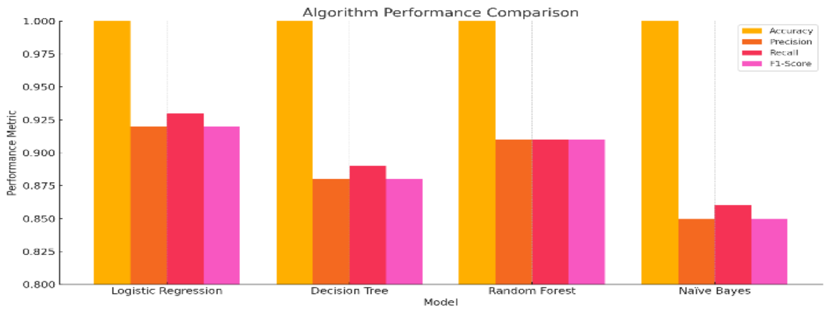

Fig. 8. Performance Comparison.

5 Conclusion

The work in this project has proven that an automatic system of detecting suicidal danger-level depression on the basis of combined text-audio input is feasible and highly accurate. A system is developed, based on Logistic Regression, that can identify mental states and offer an effective means of diagnosis and treatment of mental health. This kind of dual-input model, which combines textual and audio inputs, allows for a larger framework compared to single-input ones. Rapid feedback and personalized videos recommendation according to user mood to provide timely support systems for preventable interventions. System operation from a distance and in an anonymous fashion makes this system a convenient tool to improve the poor service availability of many mental health services to combat location and stigma barriers, especially at the areas with limited resources. Despite its system showing a positive performance, the development of the system has to reach the following challenges, such as Speech-to-Text (S-T-T) accuracy and inclusivity of speech patterns and the diversity of the dataset. Future works include focussing more on increasing dataset to improve generalisation, integration of the system to existing mental health service and most importantly riding in continuous emotion tracking, UI accommodating all users and content multilingual for globalisation. This approach represents an important step toward mental health care that is accessible quickly as clinical service that is personal and accessible that can potentially be integrated into conventional mental health services and offers sustainable solutions cope with global needs for mental health.

References

[1] Wang12, N., Chiong13, R., Kamil, R., Zhang, W., Al-Haddad, S. A. R., & Ibrahim, N. (2024). Depression detection using speech audio and text: A comprehensive review focusing on deep learning methods.

[2] Dhelim, S., Chen, L., Ning, H., & Nugent, C. (2023). Artificial intelligence for suicide assessment using Audiovisual Cues: a review. Artificial Intelligence Review, 56(6), 5591-5618.

[3] Madanian, S., Parry, D., Adeleye, O., Poellabauer, C., Mirza, F., Mathew, S., & Schneider, S. (2022, January). Automatic speech emotion recognition using machine learning: digital transformation of mental health. In Proceedings of the Annual Pacific Asia Conference on Information Systems (PACIS).

[4] Chen, Y., Li, J., Song, C., Zhao, Q., Tong, Y., & Fu, G. (2024). Deep Learning and Large

Language Models for Audio and Text Analysis in Predicting Suicidal Acts in Chinese Psychological Support Hotlines. arXiv preprint arXiv:2409.06164.

[5] Yadav, U., Sharma, A. K., & Patil, D. (2023). Review of automated depression detection: Social posts, audio and video, open challenges and future direction. Concurrency and Computation: Practice and Experience, 35(1), e7407.

[6] Carson, N. J., Mullin, B., Sanchez, M. J., Lu, F., Yang, K., Menezes, M., & Cook, B. L. (2019). Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. PloS one, 14(2), e0211116.

[7] Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Transactions on Computer-Human Interaction (TOCHI), 27(5), 1-53.

[8] Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Transactions on Computer-Human Interaction (TOCHI), 27(5), 1-53.

[9] Rabbany, K. G., Shoiynbek, A., Kuanyshbay, D., Mukhametzhanov, A., Bekarystankyzy, A., & Shoiynbek, T. (2024). A REVIEW ON MACHINE LEARNING APPROACHES FOR THE DETECTION OF SUICIDAL TENDENCIES. World science, (3 (85)), 7-21.

[10] Rana, R., Latif, S., Gururajan, R., Gray, A., Mackenzie, G., Humphris, G., & Dunn, J. (2019). Automated screening for distress: A perspective for the future. European journal of cancer care, 28(4), e13033.

[11] Ye, J. (2021). Advancing mental health and psychological support for health care workers using digital technologies and platforms. *JMIR formative research*, *5*(6), e22075.

[12] Oram, S., Fisher, H. L., Minnis, H., Seedat, S., Walby, S., Hegarty, K., ... & Howard, L. M. (2022). The Lancet Psychiatry Commission on intimate partner violence and mental health: advancing mental health services, research, and policy. The Lancet Psychiatry, 9(6), 487-524.

[13] Lasalvia, A., Patuzzo, S., Braun, E., & Henderson, C. (2023). Advance statements in mental healthcare: time to close the evidence to the practice gap. Epidemiology and Psychiatric Sciences, 32, e68.

[14] Reynolds 3rd, C. F., Jeste, D. V., Sachdev, P. S., & Blazer, D. G. (2022). Mental health care for older adults: recent advances and new directions in clinical practice and research. World Psychiatry, 21(3), 336-363.

[15] Schuck, J., Backes, E. P., Western, B., & Wang, E. A. (Eds.). (2021). Decarcerating correctional facilities during COVID-19: advancing health, equity, and safety.