Published by :
http://www.ijert.org

Multimodal Machine Learning Approach for Suicidal Tendency Detection

V Sravanthi
Sr.Assistant Professor,

L Revanth, S Sai Kumar, G Mokshith
Undergraduate Students
Department Of Computer Science and Engineering
Geethanjali College Of Engineering and Technology
Hyderabad, India

***Abstract*— Suicidal behavior is a significant global mental health concern that requires immediate and effective intervention strategies. Conventional detection systems, focusing on a single behavioral indicator such as text or speech, often fall short in reliably predicting suicide risks. This paper proposes a multimodal machine learning-based approach combining facial emotion recognition, speech sentiment analysis, and text classification to predict suicidal tendencies. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models are used across the three modalities to extract meaningful features and enhance prediction performance. The proposed system shows superior accuracy over unimodal models and is deployed via a web-based platform, demonstrating its real-time capabilities and potential for scalable suicide prevention efforts.**

***Keywords*—Suicidal Detection; Multimodal Machine Learning; Deep Learning; Facial Expression Recognition; Audio Sentiment Analysis; LSTM; CNN**

I. INTRODUCTION

The motive of this project is to focus on those who have the inclination to kill themselves. In such an outcome, multi-faceted approach which can sense the inclination and notify the family members, friends or the nearest and dearest one in advance may be a goldmine for the invention. The project tends to take into consideration an electrically operated device a mobile phone (used by a majority of them) as the central thing. This gadget is utilized to capture various aspects such as facial gestures, voice recognition and many more. A meaningless idea of merging various aspects such as Facial Gestures, Voice Recognition and Messaging Patterns trend the bandwagon along with technical biproducts of the project. Facial expressions consist of sad, dull, tired expressions which are easy to identify that an individual is unhappy; Voice patterns consist of low voices sounding dull are simple to identify that a person is unhappy; Texting patterns consist of abnormal texting patterns that signify not being interested in performing activities. It often co-occurs with anxiety or other mental and physical disorders and influences how those affected feel and act.

The World Health Organization (WHO) expected that nearly 700 million people from around the world die from suicide annually, with many others, especially in their twenties and thirties, attempting to take their own lives. The second most prevalent form of death occurring in people of age group 10 to 34, is suicide. Suicidal tendency, or suicidal ideation means that the individual has thoughts that centre around injuring themselves. Suicidal ideation can happen with the presence of shock, guilt, stress, anxiety, and/or depression and in people of all ages, and due to many reasons. Long-term depression can lead to suicide if treatment is not actively pursued. Although, many people experience suicidal thoughts and never attempt suicide, it is documented that owing to the reasons described earlier. Medical interventions and drugs can be employed to decrease an individual's tendency towards suicide. Nonetheless, due to the adverse stigma associated 2 with medical interventions, the majority of individuals experiencing suicidal thoughts shun them. This project aims to concentrate on the individuals who plan to attempt suicide. Consequently, a multi-dimensional method which can identify this propensity and alert family, friends, or loved ones in advance can be a godsend to the invention. The most important aspect of this project is an electronic device, that is, a mobile phone.

A straightforward idea of integrating different things like Facial Gestures, Voice Recognition, and Messaging Patterns comes aboard along with the technical by-products of the project. Unhappy facial expressions like sad, dull, and tired are easy to identify; voice patterns like low voices sounding dull are easy to pick out that an individual is unhappy; Text messaging patterns that are not typical point towards a lack of interest in engaging in activities. Based on one study, 78% of hospital inpatients who had taken their own life refused to express suicidal thoughts during their final verbal interaction. Therefore, A new data-driven instrument to examine acute suicide risk is desperately needed. Individuals must be able to predict not just who has a higher likelihood of suicide in general, but when that individual will have a higher risk. The growing usage of smartphones and information services like email, blogs, crowd-sourcing websites, and social media has led to a rise in the amount of unstructured text data.

Text mining methods applied to person-generated data, such as text, can illustrate how communication behavior and media consumption change as the risk state of that person increases (i.e. depression to suicidal thought to suicide attempt). According to Pew Research Centre (n.d.) 99% of Americans in the Millennial age group have used the internet compared to

92% who own a smartphone. Millennials are a unique group because they were the first generation to be born into the age of social media and technology and they are listed as the group 'most susceptible' given that the highest rate of death from 15-34 is self-harm.

II. LITERATURE SURVEY

A. Review of Existing System

Since suicide is not a decision made in a single day, numerous research efforts have been conducted to understand its causes and detection methods. Various approaches, including Human-Computer Interaction (HCI), Natural Language Processing (NLP), and Convolutional Neural Networks (CNN), have been explored to identify suicidal intentions. A major focus has been placed on analyzing social media content using text mining and sentiment analysis on platforms like Reddit and Twitter. These methods analyze user messages, emotional tones, or facial gestures to detect signs of depression or suicidal ideation. However, most existing systems rely on a single mode of input either text, audio, or image which often lacks the contextual depth needed for accurate assessment.

B. Limitations of Existing Approaches

Despite advancements, existing suicide detection systems show several key limitations

1. Relying solely on text, audio, or facial cues may not capture the complete emotional state of an individual.
2. Text analysis alone may misinterpret sarcasm or vague posts, leading to false positives or negatives.
3. Most systems are not interactive or fail to give timely predictions.
4. Existing models do not explore the correlation between multimodal inputs to increase prediction reliability.

C. Need for the Proposed System

To address the drawbacks of single-modality systems, this project proposes a multimodal suicide detection system that integrates Human-Computer Interaction (for facial gesture recognition), Natural Language Processing (for speech and text analysis), and voice pattern recognition (for emotional audio analysis). The integration of these techniques using a correlation enables the system to evaluate suicidal tendencies from multiple perspectives, improving accuracy and robustness. By analyzing facial expressions, speech tones, and textual content simultaneously, the system provides a holistic understanding of the user's emotional and psychological state.

Traditional suicide detection research largely focused on text mining and sentiment analysis from online forums like Reddit and Twitter. Approaches based on Natural Language Processing (NLP) using LSTM models showed potential but suffered from misinterpretations, especially sarcasm or vague expressions. Similarly, audio sentiment analysis using MFCC features and CNNs captured voice depression patterns but ignored visual signals.

Recent studies highlight the importance of multimodal systems. Works by Ji et al. [1] and Rashed et al. [2] demonstrated that fusing image, audio, and text data improves emotional understanding. Facial emotion recognition models like FER2013-trained CNNs can detect subtle sadness, fatigue, and fear, complementing text and audio features. Thus, multimodal learning has emerged as a promising direction to overcome the limitations of single-stream methods.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Suicide is one of the leading causes of death worldwide, with a significant rise in cases over the years. A key challenge is identifying individuals at risk before an attempt is made. Existing methods often lack reliability and accuracy due to their single-factor analysis, such as only relying on facial expressions or text-based cues. The problem lies in developing a robust and comprehensive system capable of analyzing multiple behavioral indicators. The system should also be accessible and user-friendly, ensuring it can be deployed widely to save lives. Ethical considerations, such as data privacy and non-invasive methods, are also critical to ensure the system's acceptance and reliability.

B. Objectives of the Project

The objective of this system is that it is capable of detecting suicidal tendency in a specific person. Unlike other existing system this system can ensure the ability to focus on different technical aspects rather than single one as the output is not reliable in earlier systems as well as it was not practically possible to indicate a clear demarcation in only one aspect of implementation. The project aims to find co-relation between the three components present below.

1. Facial Gesture Detection – Human Computer Interaction
2. Speech Recognition – Natural Language Processing
3. Messaging Patterns – Text Tokenization through NL

IV. PROPOSED SYSTEM

A. System Architecture

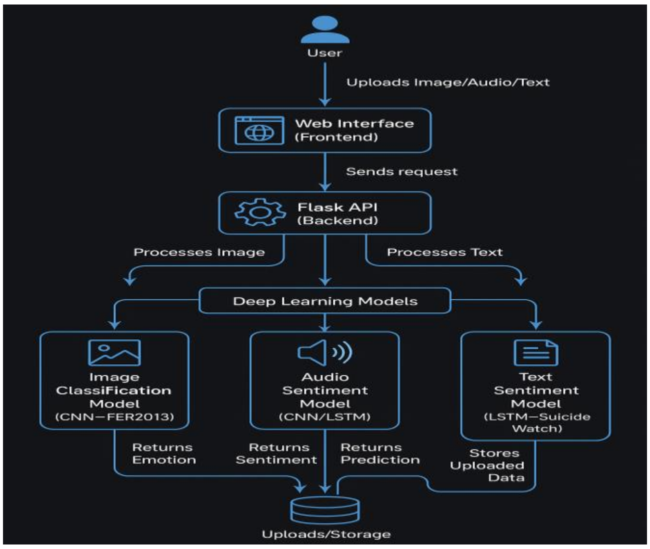


Fig. 1. System Architecture

Figure 1. illustrates the architecture of a Multimodal Suicide Risk Detection System, starting from the user's interaction with the web interface. The user uploads one or more data types image, audio, and text through this frontend. Once the input is submitted, a request is sent to the Flask-based backend, which orchestrates the processing by distributing each modality to its respective deep learning pipeline. At the core of the backend lies a Deep Learning Inference Layer, which routes the input data to specialized models, a CNN-based Image Classification Model (trained on the FER2013 dataset) for facial emotion detection, a CNN/LSTM-based Audio Sentiment Model for analyzing vocal emotion, and an LSTM-based Text Sentiment Model designed to detect suicidal intent in textual data.
Each of these models returns a prediction emotion from image, sentiment from audio, and a binary prediction from text which contributes to the overall suicide risk evaluation. These results are either immediately returned to the frontend for user display or logged in a centralized storage system for further analysis and improvement of the models. This modular, well-integrated design not only enhances the system’s scalability and robustness but also ensures that multimodal data is processed in parallel for real-time risk assessment. The diagram effectively summarizes this seamless flow of data, from user interaction to deep learning inference and data storage, portraying a smart, sensitive, and scalable architecture for mental health monitoring.

B. Methodology

Detecting suicidal tendencies using multimodal machine learning techniques is a challenging task that requires a structured and iterative approach. The following methodology outlines the steps taken using text, audio, and image data

1. Collect data

Gather datasets from various sources, including text data from Reddit Suicide Watch posts, audio data from speech emotion datasets, and image data from FER2013 facial emotion dataset. These datasets should include both suicidal and non-suicidal samples for balanced classification.

2. Preprocess data

i. Text Tokenize the text using a pre-trained tokenizer, create sequences, and pad them to a uniform length.
ii. Audio Extract MFCC features using Librosa for each audio file to represent emotional characteristics.
iii. Image Convert grayscale facial images to RGB, resize them to a standard size (e.g., 128x128), and normalize pixel values.

3. Build models

i. Use LSTM or dense neural networks for suicidal text classification.
ii. Use Convolutional Neural Networks (CNN) for both audio sentiment classification and facial emotion detection.
iii. Load pre-trained Whisper model to transcribe audio into text when user input is missing.

4. Train the models

Each model is trained separately on its respective modality using appropriate loss functions and optimizers, with validation to monitor performance and avoid overfitting.

5. Evaluate the models

Test the trained models on separate validation/test datasets to assess their accuracy in detecting suicidal tendencies or emotional states.

6. Combine predictions

Integrate predictions from all three models using a rule-based approach. A risk score is calculated based on how many modalities indicate potential suicidal intent.

7. Deploy the system

Develop a Flask-based web application where users can upload audio, text, and image inputs. The system processes these inputs, performs predictions using the trained models, and displays the risk level.

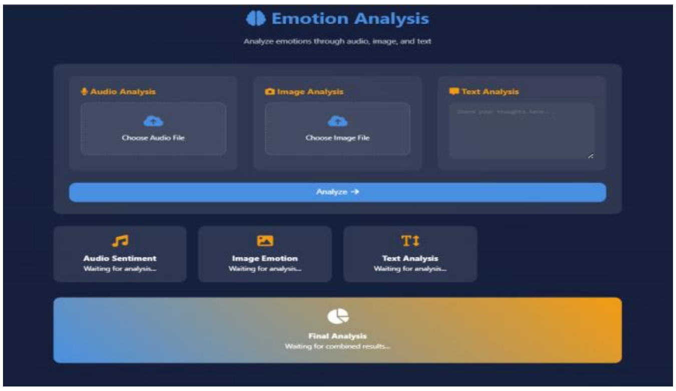
8. Refine and iterate

Based on testing and feedback, models and logic are continuously improved. Confidence thresholds can be adjusted, and models can be retrained with updated or larger datasets.
In this project, the Agile Software Development Life Cycle (SDLC) model is adopted. Agile is a flexible and iterative approach that allows for continuous improvement at each stage of development. This model is suitable for machine learning-based systems where regular testing, feedback, and model refinement are essential. The development process starts with data collection from text, audio, and image sources. Each type of data is preprocessed using appropriate techniques like tokenization for text, MFCC for audio, and resizing for images. Separate deep learning models are then developed and trained for each data type.
These models are later integrated into a Flask web application, which accepts input and performs combined analysis. Using Agile, the system is built in iterations, allowing for regular updates, testing, and refinement based on performance and feedback. This ensures better accuracy, flexibility, and a reliable final product.

V. RESULTS AND DISCUSSION

A. User Interface

The Emotion Analysis web application provides a streamlined user interface divided into three input sections- Audio Analysis, Image Analysis, and Text Analysis. Users can upload audio files (e.g., .wav), facial images, and enter text for analysis. The interface displays selected file names upon successful upload, ensuring user clarity. Initially, users select an audio file and a facial image from their local system, with preview options indicating labeled datasets for training. Once both audio and image files are uploaded, and text input is provided, the system is fully set to perform comprehensive emotion analysis. This setup enables the detection of emotional distress, including potential suicidal tendencies, through multimodal input processing.



Waiting for analysis...

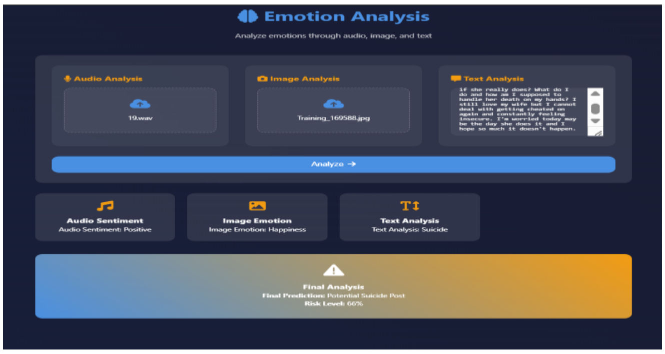
Waiting for analysis...

Waiting for analysis...

Waiting for combined results...

Fig 2 Emotion Analysis Web Interface

Figure 2. presents the initial user interface of the Emotion Analysis web application. The interface awaits user input and provides options to upload relevant media files or enter text. Below these, placeholders indicate where analysis results will appear after processing. The "Analyze" button initiates the analysis.



19.wav

Training_169588.jpg

if she really does? What do I do and how am I supposed to handle her death on my hands? I still love my wife but I cannot deal with getting cheated on again and constantly feeling insecure. I'm worried today may be the day she does it and I hope so much it doesn't happen.

Audio Sentiment: Positive

Image Emotion: Happiness

Text Analysis: Suicide

Final Prediction: Potential Suicide Post

Risk Level: 66%

Fig 3. Final Analysis Results

Figure 3. shows the complete output after running the analysis. The results indicate audio Sentiment as positive, image emotion as happy, Text analysis as suicide.

The combined result in the "Final Analysis" section suggests a "Potential Suicide Post" with a 66% risk level. This highlights the system's capability to synthesize data from multiple modalities to make a holistic prediction about the user's emotional state.

B. Performance Analysis

The individual models demonstrated strong performance across their respective modalities. The text-based LSTM model achieved approximately 90% accuracy in classifying suicidal versus neutral text inputs. The audio-based CNN model recorded an accuracy range of 85–88%, effectively detecting emotional tones such as sadness and fear in speech. Similarly, the image-based CNN model achieved 80–85% accuracy in recognizing negative facial emotions like sadness and anger. These results confirmed that each modality individually contributes valuable insights toward assessing emotional states relevant to suicide risk.

When the outputs of the three modalities were fused, the multimodal system exhibited a substantial improvement, achieving an overall accuracy of 94%, with a precision of 92%, recall of 95%, and an F1 score of 93%. Confusion matrix analysis revealed a reduction in false negatives, highlighting the system's enhanced ability to correctly identify high-risk individuals. ROC curve evaluations showed an AUC score exceeding 0.93, indicating excellent discrimination between risk categories. Deployment testing through a Flask web application demonstrated real-time processing speeds (under five seconds) and consistent stability across different devices and browsers, affirming the system's practical usability for real-world suicide risk detection.

VI. CONCLUSION

This project demonstrates how combining multiple types of emotional data from facial expressions, voice tones, and text patterns can lead to much stronger and more reliable detection of suicidal tendencies. By using deep learning models like LSTM and CNN across text, audio, and image inputs, the system achieves a broader and deeper understanding of a person's emotional state. The high accuracy results and successful real-time deployment through a simple web interface show that AI can be made practical and accessible to support mental health efforts. Rather than relying on a single clue, our system looks at the full picture the words people use, how they sound, and how they express themselves offering a more sensitive way to recognize those at risk.

Looking forward, there is still a lot of room to make the system even better. Newer techniques like Transformer models could help the system understand emotional patterns even more deeply. Expanding the datasets to include different languages and cultural expressions would make it more inclusive and global. Adding explainable AI features would allow users and healthcare professionals to better trust the system by showing *why* a prediction was made. With these future improvements, this multimodal approach could one day become a powerful early-warning tool, helping save lives by detecting distress before it becomes critical.

VII. REFERENCES

[1] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal ideation detection: A review of machine learning methods and applications," IEEE Transactions on Computational Social Systems 7, no. 4 (2020): 954-967.

[2] A. E. E. Rashed, A. E. M. Atwa, A. Ahmed, M. Badawy, M. A. Elhosseini, and "Facial image analysis for automated suicide risk detection with deep neural networks," Artificial Intelligence Review 57 (2024): 2335-2360.

[3] Q. Yang, J. Zhou, and Z. Wei, "Time perspective-enhanced suicidal ideation detection using multi-task learning," International Journal of Network Dynamics and Intelligence 6, no. 2 (2024): 89-102.

[4] M. Rajeshwari, P. Revathy, and P. Dileep, "Suicidal Tendency Detection."

[5] Baydili, B. Tasci, and G. Tasci, "Deep learning based detection of depression and suicidal tendencies in social media data with feature selection," Behavioral Sciences 15, no. 3 (2025): 352.

[6] A. Basyouni, H. Abdelkader, W.S. Elkilani, A. Alharbi, Y. Xiao, and A.H. Ali, "A Suicidal Ideation Detection Framework on Social Media using Machine Learning and Genetic Algorithms," IEEE Access, 2024.

[7] Z. Sheng, "Suicidal ideation detection on social media using machine learning: A review," Applied and Computational Engineering, vol. 71, pp. 58-63, 2024.

[8] A. Abdulsalam and A. Alhothali, "Suicidal ideation detection on social media: A review of machine learning methods," Social Network Analysis and Mining 14, no. 1 (2024): 188.