

RESEARCH ARTICLE

OPEN ACCESS

Detecting Trauma Symptoms and Connecting Doctors Using Voice and Text Analysis

Suda Lagunada Sai Venkata Trinesh
Dept. of Computer Science & Engineering
(AIML)
Parul Institute of Engineering and
Technology
Parul University
Vadodara, India
Enrollment No. 2303031460167

Shaik Abraar
Dept. of Computer Science & Engineering
(AIML)
Parul Institute of Engineering and
Technology
Parul University
Vadodara, India
Enrollment No. 2303031460162

Sammiti Bharath Kumar
Dept. of Computer Science & Engineering
(AIML)
Parul Institute of Engineering and
Technology
Parul University
Vadodara, India
Enrollment No. 2303031460158

Irala Venkata Durga Prasad
Dept. of Computer Science & Engineering
(AIML)
Parul Institute of Engineering and
Technology
Parul University
Vadodara, India
Enrollment No. 2303031460062

Parsuboina Naveen
Dept. of Computer Science & Engineering
(AIML)
Parul Institute of Engineering and
Technology
Parul University
Vadodara, India
Enrollment No. 2303031460146

Abstract - Psychological trauma often remains unnoticed because it is subtle and rarely monitored continuously, and manual assessment is time-consuming and subjective. This paper describes the design of an AI-based web application that analyzes user-supplied voice and text for possible trauma-related indicators. Text is analyzed with Natural Language Processing (NLP) techniques, while voice is processed with speech analysis to extract emotional characteristics such as tone, pitch and speech patterns. The two sources are combined in a multimodal approach, and the user's state is classified as Low, Medium or High risk. When high risk is identified, the system lets the user contact a healthcare professional by call or messaging. The system also includes registration and login, report generation, database storage and an administrative module. The aim is early awareness, accessibility and timely support. The project report documents the design, requirements and implementation approach; it does not report measured accuracy or clinical validation. The system is a decision-support tool and is not intended to replace professional diagnosis.

Keywords - mental health, multimodal analysis, natural language processing, risk classification, trauma detection

1. Introduction

Artificial intelligence is increasingly applied to healthcare, including mental-health monitoring. Psychological trauma is difficult to identify early: it often lacks continuous observation, and conventional assessment depends on manual evaluation and self-reporting, which are slow and subjective [1,3].

Text and voice are both practical sources of emotional information. Language use reflects psychological state [6], and NLP has been applied to mental-health text [1,5]. Speech carries tone, pitch, rhythm and intensity that relate to conditions such as depression and post-traumatic stress disorder (PTSD) [3,4,7]. Text conveys meaning but little vocal tone, whereas speech conveys tone but not necessarily content, which motivates combining them.

This paper presents a web application that analyzes voice and text to estimate trauma risk and connects high-risk users with healthcare professionals. The contributions, as described in the project report, are: (1) a voice-and-text trauma-analysis workflow; (2) NLP-based analysis of text; (3) speech-based

emotional feature extraction (for example, MFCC); (4) multimodal fusion of both outputs; (5) Low, Medium and High risk classification; (6) doctor connectivity for high-risk cases; and (7) a web architecture with database, administration and security requirements.

2. Related Work

The project report reviews 25 studies. Because several entries are incomplete, they are synthesized here by theme, with bibliographic details as listed in the report.

2.1. Text-Based Mental Health Detection

Calvo and Milne review NLP for mental-health applications using non-clinical texts [1], and Coppersmith and Dredze apply NLP to detect depression and trauma in online text [18]. Larsen and Boonstra report that RNN/LSTM models with Word2Vec or GloVe embeddings outperform keyword and Naïve Bayes/SVM baselines at capturing context [8]. Ji and Pan describe MentalBERT, a pretrained model fine-tuned on mental-health data [14], building on BERT [5]. Linguistic markers of psychological state were studied earlier by Pennebaker et al. [6].

2.2. Speech Emotion Recognition

Cummins et al. review speech analysis for depression and suicide-risk assessment [3], and Low et al. examine acoustic descriptors for depression detection [4]. Scherer et al. study voice quality as an indicator of depression and PTSD [7]. Schuller and Steidl [9] and Neumann and Vu [17] report CNN/LSTM or CNN/RNN approaches using MFCC, pitch and spectral features. Handcrafted-feature methods are reported to generalize poorly across speakers and noisy conditions.

2.3. Multimodal Analysis

Poria and Cambria [10], Zeng and Pantic [19] and Al Hanai et al. [13] combine speech and text using feature-level or decision-level fusion, and the report states that these studies found multimodal systems to outperform unimodal ones. The report also lists the DAIC-WOZ corpus [15] as a multimodal depression dataset.

2.4. Digital Mental Health, Chatbots and Stress Detection

Miner et al. examine how people discuss personal mental-health problems with machines [2]. Fitzpatrick and Darcy [11] and Torous and Wisniewski [16] discuss conversational agents and digital interventions that address accessibility, cost and stigma. Kumar and Singh [12] and Pan and Huang [20] address stress and mental-health monitoring from speech, text and behavioral data.

2.5. Doctor Recommendation

Sharma and Gupta [21] and Kumar and Verma [22] describe machine-learning systems that match users to professionals by specialization, severity and availability, with alerts for high-risk cases.

2.6. Research Gap

The reviewed studies indicate that most work addresses one modality or one stage (detection or recommendation) separately. The project addresses this gap by combining voice analysis, text analysis, multimodal processing, risk classification and healthcare-professional connection in one web application. No claim of superiority over prior work is made, as no comparative results are available.

Table 1. Summary of related work (as listed in the report)

Theme	Representative work	Method reported
Text	[1],[8],[14],[18]	NLP, LSTM, BERT-type models
Speech	[3],[4],[9],[17]	MFCC, pitch; CNN/LSTM/RNN
Multimodal	[10],[13],[19]	Feature/decision-level fusion
Digital health	[2],[11],[16]	Chatbots, telemedicine
Doctor matching	[21],[22]	ML matching, severity alerts

3. Materials and Methods

3.1. Problem Statement and Objectives

Trauma-related distress frequently goes undetected because of limited continuous monitoring and reliance on manual, subjective assessment, which delays professional intervention. Existing solutions were observed in the project to be limited by single-modality analysis and absence of real-time doctor connectivity.

The project aims to (i) analyze inputs with NLP and speech processing; (ii) improve detection accuracy through a multimodal approach (a design aim, not a measured outcome); (iii) classify risk as Low, Medium or High; (iv) provide real-time analysis and feedback; and (v) enable communication with healthcare professionals. The project focuses on early detection and support rather than diagnosis.

3.2. Proposed System

The pipeline is: user registration and login, voice or text input, preprocessing, feature extraction, AI analysis, multimodal fusion, risk classification (Low, Medium or High), analysis report, and doctor connection if risk is high; otherwise monitoring continues. Users can view results and track their status over time. Table 2 summarizes the modules.

Table 2. System modules and functions

Module	Function
Input	Captures voice and text via web interface
Preprocessing	Text cleaning, audio noise removal
Feature extraction	MFCC (voice); sentiment, embeddings (text)
Prediction	Fusion and Low/Medium/High classification
Alert / doctor	Notifies doctors; call or messaging
UI	Results, reports, recommendations
Admin / database	Manages users, doctors, data

Per the software requirements specification (SRS), doctor recommendation is triggered only for high-risk cases to avoid unnecessary alerts, and outputs are advisory rather than diagnostic.

3.3. System Architecture

The report describes six layers: a presentation layer (HTML/CSS web interface); an application layer (Flask backend handling API requests, preprocessing, model execution and responses); an AI model layer (BERT, TF-IDF and LSTM for text; CNN/BiLSTM for speech emotion); a fusion layer combining model outputs; a data layer (relational database for inputs, predictions and logs); and a doctor connectivity module triggering alerts and telemedicine contact.

3.4. Processing Workflow and Data Flow

Figure 1 gives the flowchart: sign-up and login, dashboard, text or voice input, preprocessing, feature extraction, AI analysis, fusion of text and voice results, and an "Is risk

high?" decision leading to doctor connection or continued monitoring. Optional features include doctor search and password reset.

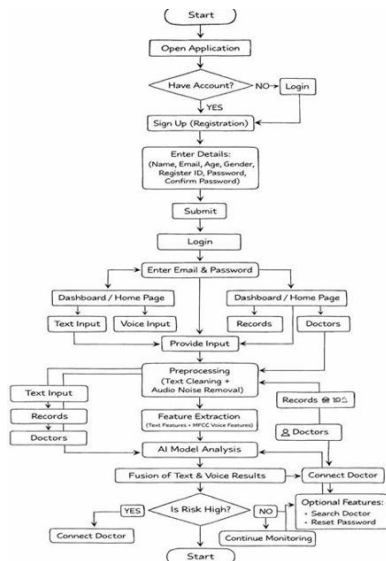


Fig. 1 System flowchart of the proposed trauma detection system.

3.5. Use Case Modeling

Three actors are defined. Users register, submit input, view reports and contact doctors. Doctors view patient reports and respond by call or message without modifying system data. Administrators manage users, doctors and the database. **Figure 2** shows the UML use case diagram. The administrator role is described in the text only. No separate data flow, sequence or activity diagrams are present in the report.

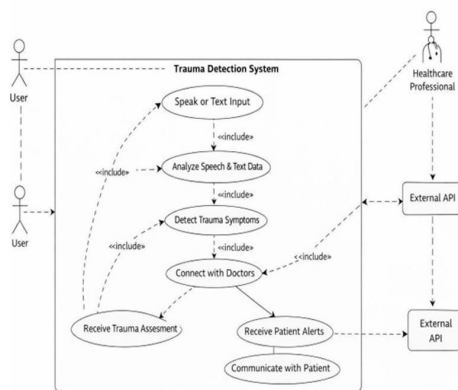


Fig. 2 UML use case diagram of the trauma detection system.

3.6. Development Methodology

The Waterfall model was chosen: requirements, design, implementation, testing, deployment and maintenance, completed sequentially. Requirements were gathered from the perspectives of users, doctors and administrators.

3.7. Implementation

Frontend: HTML and CSS (JavaScript is also listed). **Backend:** Python with Flask, handling APIs and model execution; communication is HTTP-based and may be extended with WebSockets. **NLP:** BERT, TF-IDF and LSTM are listed; sentiment and embeddings serve as features. **Speech:** MFCC features with Librosa/PyAudio listed; CNN/BiLSTM models for emotion recognition. **AI/ML frameworks:** TensorFlow, Keras and PyTorch are listed as possibilities, with scikit-learn for preprocessing; the final choice is not explicitly specified. **Database:** SQLite (development) and MySQL (production), with PostgreSQL also named in the SRS.

Security: the SRS specifies HTTPS, encryption in transit and at rest, hashed passwords, role-based access control (RBAC), multi-factor authentication (MFA), logging and vulnerability management as requirements; MFA and stronger encryption are also listed as future enhancements, so they should not be assumed implemented. OAuth-based API authentication is listed. **Doctor communication:** call or messaging/telemedicine (chat or video) after a high-risk alert. Cloud platforms (AWS, GCP, Azure) or local servers are named as deployment options; actual deployment is not confirmed.

Table 3. Key requirements

Type	Requirement (SRS)
Functional	Voice/text input, preprocessing, feature extraction, risk prediction, report, doctor connection
Performance	Fast response, multi-user scalability (no values given)
Security	RBAC, encryption, hashing, MFA, audit logs
Legal/ethical	Consent, GDPR or local privacy law, non-diagnostic disclaimer

4. Results and Discussion

4.1. Results and Analysis

The project report primarily documents the system design, implementation approach and intended testing process. It does not provide sufficient measured experimental results (dataset size, accuracy, precision, recall, F1-score, confusion matrix or processing time) to make quantitative performance claims, and none are made here. Statements in the report that multimodal analysis improves accuracy are design expectations supported by the literature [10,13,19], not results of this project. The design-level outcome is a specified architecture, module set and workflow. Datasets named in the SRS (DAIC-WOZ [15], IEMOCAP, CLPpsych) are listed as references for training and evaluation; their actual use is not documented.

4.2. Testing and Validation

Table 4. Testing strategy (planned/described; no results reported)

Type	Purpose
Unit	Individual modules
Integration	Combined modules
System	Complete system
User	Usability
Performance / Security	Response time; data protection
Model evaluation	Precision, recall, F1-score (intended)

The report states that errors were identified and fixed before deployment, but gives no test cases or pass/fail outcomes, so none are reproduced.

4.3. Discussion

Text supplies semantic content and speech supplies emotional tone, so combining them may give a fuller picture than either alone, as the multimodal literature suggests [10,19]. Tiered risk levels allow a graded response, and restricting doctor alerts to high risk limits unnecessary escalation. Voice recordings and text are sensitive, so consent, encryption and privacy compliance are central requirements. Outputs must be read as advisory insights, and no clinical effectiveness is claimed.

4.4. Limitations

The report identifies that accuracy depends on the quality and availability of training data, that deep models require substantial computation and may perform poorly on low-end devices, and that sensitive-data handling must meet ethical and privacy regulations. No measured evaluation or clinical validation exists. Several survey entries have incomplete bibliographic data and need verification.

4.5. Future Work

The report lists: advanced transformer and multimodal deep-learning models; Android/iOS applications; multilingual support; real-time monitoring via wearables or voice tracking; Explainable AI to assist doctors; and multi-factor authentication with stronger encryption.

5. Conclusion

This paper described an AI-based web application that analyzes voice and text with NLP and speech processing, fuses them for Low, Medium and High risk classification, and connects high-risk users with healthcare professionals. The contribution is an integrated design, with requirements and architecture, for early awareness and accessible support. Quantitative evaluation and clinical validation remain future work, and the system is not a replacement for professional diagnosis.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors received no specific funding for this work. [Edit if applicable.]

Acknowledgments

The authors thank the Department of CSE (AIML), Parul University, and their project guides.

References

- [1] R. A. Calvo and D. N. Milne, "Natural language processing in mental health applications using non-clinical texts," *Nat. Lang. Eng.* vol. 22, no. 5, pp. 649–685, 2016.
- [2] A. S. Miner, A. Milstein, J. T. Hancock *et al.*, "Talking to machines about personal mental health problems," *JAMA*, vol. 318, no. 13, pp. 1217–1218, 2017.
- [3] N. Cummins, S. Scherer, J. Krajewski *et al.*, "A review of depression and suicide risk assessment using speech analysis," *Speech Commun.*, vol. 71, pp. 10–49, 2015.
- [4] L. A. Low, N. C. Maddage, M. Lech *et al.*, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 574–586, 2011.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019.
- [6] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use," *Annu. Rev. Psychol.*, vol. 54, pp. 547–577, 2003.
- [7] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency, "Investigating voice quality as a speaker-independent indicator of depression and PTSD," in *Proc. Interspeech*, 2013.
- [8] M. E. Larsen and T. W. Boonstra, "Detecting depression from social media text using deep learning," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 1–10, 2018.
- [9] B. Schuller and S. Steidl, "Speech emotion recognition using deep learning techniques," *IEEE Trans. Affective Comput.*, vol. 10, no. 4, pp. 1–12, 2019.
- [10] S. Poria and E. Cambria, "Multimodal emotion recognition using deep learning," *ACM Trans. Multimedia Comput.*, vol. 15, no. 1, pp. 1–19, 2019.
- [11] K. Fitzpatrick and A. Darcy, "Conversational agents for mental health applications," *J. Med. Internet Res.*, vol. 19, no. 2, pp. 1–10, 2017.
- [12] A. Kumar and D. Singh, "AI-based web application for real-time mental health monitoring using speech and text analytics," *Int. J. Med. Inform.*, vol. 169, Art. 104933, 2023.

- [13] T. Al Hanai, M. Ghassemi, and J. Glass, “Multimodal depression detection using deep learning,” in *Proc. Interspeech*, 2018, pp. 1–5.
- [14] S. Ji and X. Pan, “MentalBERT: Pretrained language models for mental healthcare,” *ACL Findings*, 2021.
- [15] J. Gratch *et al.*, “The Distress Analysis Interview Corpus (DAIC-WOZ): A dataset for depression detection,” *World Psychiatry*, vol. 10, no. 3, 2019.
- [16] J. Torous and M. Wisniewski, “Digital mental health interventions: Opportunities and challenges,” *World Psychiatry*, vol. 19, no. 3, pp. 1–12, 2020.
- [17] M. Neumann and N. T. Vu, “Speech emotion recognition for trauma detection using deep learning,” *IEEE Trans. Affective Comput.*, vol. 9, no. 2, pp. 1–12, 2018.
- [18] G. Coppersmith and M. Dredze, “Natural language processing for mental health detection,” in *Proc. NAACL-HLT*, 2015, pp. 1–10.
- [19] Z. Zeng and M. Pantic, “Multimodal emotion recognition using speech and text data,” *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 1–15, 2019.
- [20] S. Pan and J. Huang, “Deep learning-based stress detection using behavioral data,” *IEEE Access*, vol. 8, pp. 1–12, 2020.
- [21] R. Sharma and P. Gupta, “AI-based doctor recommendation system for mental health support,” *Int. J. Healthcare Inform.*, vol. 12, no. 2, pp. 45–60, 2022.
- [22] A. Kumar and N. Verma, “AI-based doctor recommendation system for trauma care,” *Int. J. Smart Healthcare*, vol. 3, no. 2, pp. 41–50, 2024.