



# MANASAI: AN AI -POWERED EMOTIONAL AND MENTAL HEALTH DETECTION SYSTEM

**B. VIJAYENDRA GOWTHAM<sup>1</sup>, T. MURALIDHAR<sup>2</sup> and R. V V S SANDEEP NAIDU<sup>3</sup>**

**Dr. S AKILA<sup>4</sup>, Dr. K S RAMANUJAM<sup>5</sup>, Dr. T V ANANTHAN<sup>6</sup>**

**Department of Computer Science and Engineering (Artificial Intelligence),**

**Dr. M.G.R. Educational and Research Institute, Chennai 600095, India**

**Corresponding author: B. Vijayendra Gowtham**

**This work was supported in part by the Department of Computer Science and Engineering  
(Artificial Intelligence),**

**Dr. M.G.R. Educational and Research Institute, Chennai.**

## Abstract

Mental health is a critical global challenge, with escalating cases of stress and psychological disorders. Traditional support systems, while effective, face significant bottlenecks, including high consultation costs, limited accessibility, and persistent social stigma. To address these issues, **ManasAI** is developed as an AI-powered mental health chatbot providing real-time emotional support and well-being guidance. The system utilizes advanced Natural Language Processing (NLP) and sentiment analysis to identify underlying emotions in text conversations and generate tailored, empathetic responses. Uniquely, ManasAI enhances its multimodal emotion detection by integrating handwriting-based psychological pattern recognition. The technical architecture employs **FastAPI** for an efficient and scalable backend, **MongoDB** for secure data storage, and a fusion model that combines

Transformer-based architectures (for text), Convolutional Neural Networks (CNNs) (for handwriting), and anomaly detection models. ManasAI extends beyond conversation, offering mood trend visualization, risk scoring, and safety escalation protocols to proactively detect critical conditions like suicidal ideation or severe distress.

**Keywords:** Mental Health, Chatbot, NLP, Sentiment Analysis, Pattern Recognition, Handwriting Analysis, Deep Learning, Anomaly Detection, Emotion Recognition, AI in Healthcare.

## Introduction

### Overview

Good health encompasses complete physical, mental, and social well-being, with emotional awareness playing a fundamental role. Affective Computing, the field that bridges psychology and computer science, enables computational methods to understand human emotions through various modalities, including speech, handwriting, and text. ManasAI is an AI-powered platform designed to recognize and analyze human emotions across multiple input forms, leveraging NLP for textual sentiment, CNNs for handwriting analysis, and deep learning models for speech cues. This system aims to support emotional self-awareness in real time, aligning with the United Nations Sustainable Development Goal 3 (Good Health and Well-Being).

### Problem Statement and Motivation

Mental health challenges are rising due to factors like academic pressure and digital isolation. Early identification of emotional changes can drastically reduce the severity of psychological disorders, yet a simple, continuous self-monitoring tool is often lacking.

The problem ManasAI addresses is the need for a unified, AI-driven system capable of: capturing diverse inputs (text, speech, handwriting), classifying emotions (happiness, sadness, anger, fear, disgust, neutrality) with measurable confidence, providing meaningful visual feedback, and operating as an accessible, lightweight web application. The motivation is to democratize mental-health technology, making emotion-awareness tools freely available and non-judgmental.

### Objectives and Scope

#### Project Objectives

The overarching goal is to design and implement an intelligent platform that detects, analyzes, and interprets human emotions from multimodal data sources. Specific objectives include:

1. Designing a unified framework that integrates text, speech, and handwriting emotion recognition within a single web platform.
2. Developing AI models capable of learning complex emotional representations using

deep-learning techniques such as CNNs, RNNs, and transformer-based networks.

3. Implementing a FastAPI-based backend that processes incoming data, performs inference, and
4. returns emotion results in real time.
5. Visualizing emotion trends and confidence levels for users to track their emotional variations over time.
6. Promoting mental-health awareness by providing accessible, privacy-preserving tools for emotional self-evaluation.
7. Ensuring scalability and extensibility for future modules like face-emotion detection and real-time mood tracking from wearable sensors.

### Project Scope

The scope covers all stages from emotion acquisition to feedback delivery. The system focuses on emotion detection and visualization, complementing professional counseling by offering continuous emotional tracking, but is explicitly not intended for medical diagnosis or treatment. The project embodies the spirit of ethical AI, ensuring transparency in emotion prediction, secure storage, and user data privacy. The functional scope is summarized below:

Component	Scope Description
<b>Input Data</b>	Text messages, recorded speech (.wav), scanned handwriting images
<b>Processing Layer</b>	Feature extraction, signal processing, NLP pipelines, CNN-based image analysis
<b>Model Layer</b>	Training, validation, and fusion of independent emotion models
<b>Output Layer</b>	Real-time emotion classification, visualization graphs, and mood summaries

### Significance of the Study

Academically, the research demonstrates how multi-modal data fusion enhances AI accuracy in affective computing. Socially, it provides an early-

warning mechanism for emotional distress, supporting wellness initiatives under SDG Goal 3. Technologically, it integrates various AI subfields—NLP, speech analysis, and computer vision—into a single deployable framework.

## Literature Survey

### Text-Based Emotion Recognition

Text-based emotion recognition uses NLP models to interpret emotional tone from written language, commonly used in chatbots and social media monitoring.

- **Traditional Methods:** Earlier lexicon-based analysis methods used predefined emotional dictionaries but were static and failed to understand contextual emotions like sarcasm or irony.
- **Machine Learning Approaches:** Models like Naive Bayes and SVM combined with TF-IDF improved accuracy but still lacked the ability to model complex word relationships.
- **Deep Learning and Transformer Models:** The development of RNNs, LSTMs, and especially **Transformers (BERT, RoBERTa, GPT)**, revolutionized the field. These models capture long-range dependencies and contextual semantics, achieving up to 90% accuracy on benchmark datasets. ManasAI adopts similar transformer techniques for its textual emotion analysis.

### Speech-Based Emotion Recognition

Speech conveys emotional information through features like tone, pitch, rhythm, and intensity, requiring the analysis of both spectral and temporal features.

- **Feature Extraction:** Commonly used features include MFCC (Mel-Frequency Cepstral Coefficients), Chroma Features, and Spectral Centroid, which capture pitch/tone, harmonic content, and energy distribution, respectively. These are extracted using libraries like Librosa
- **Model Approaches:** Deep Models, such as CNNs and RNNs, replaced classical methods like Gaussian Mixture Models (GMM). The current state-of-the-art utilizes **Hybrid CNN-LSTM architectures** to combine spatial

(frequency patterns) and temporal (tone changes) features for improved accuracy.

- **Performance:** Deep CNN-based models achieve 85-90% accuracy on datasets like RAVDESS. ManasAI leverages similar feature extraction with lightweight CNNs to process audio locally via the FastAPI backend.

## System Architecture and Methodology

### Handwriting-Based Emotion Recognition

Handwriting carries subconscious patterns that reflect emotional states like anxiety or stress, a study known as Graphology.

- **Deep Learning Approaches:** Convolutional Neural Networks (CNNs) have significantly improved handwriting emotion recognition by automatically learning spatial hierarchies of patterns, eliminating the need for manual feature extraction. ManasAI utilizes CNNs pre-trained on datasets such as IAM and custom emotion-labelled samples. Popular CNN architectures include LeNet-5, VGG-16/VGG-19, and ResNet-50.

### Multimodal Emotion Recognition and Fusion

Integrating multiple signals (text, audio, handwriting) improves robustness and accuracy, helping to resolve ambiguous emotions.

- **Fusion Techniques:**

1. **Early Fusion:** Combining features from different modalities before the model.
2. **Late Fusion:** Merging decisions (e.g., weighted voting) from separate models.
3. **Hybrid Fusion:** Combining both early and late methods.

ManasAI utilizes a modular multimodal pipeline where models run independently, and the final emotion output is computed using a **weighted average of confidence scores** (a form of Late Fusion). This approach addresses limitations of existing systems, such as lack of text input, high computational cost, or lack of web-deploy ability.

## System Architecture

The architecture integrates a frontend, a high-performance backend, modular AI services, and a persistent database layer, prioritizing scalability, performance, and modularity.

8. **User Interface (UI):** Built with HTML, CSS, and JavaScript, it collects multimodal inputs (Text, Voice, Image).
9. **FastAPI Backend:** Acts as the bridge, handling API requests, model routing, and sending real-time JSON responses.
10. **Emotion Analysis Modules:** Separate modules for Chat (Text Model), Voice (Audio Model - CNN-LSTM), and Handwriting (CNN Model) perform independent inference.
11. **Fusion Layer/Mood Analyzer:** Combines the emotion results and aggregates data to show long-term mood patterns.
12. **MongoDB Database:** Stores emotion records, timestamps, and model predictions.
13. **Visualization Layer:** Displays the final emotion results via graphs, percentages, and reports.

## System Requirements and Implementation

### Hardware and Software Requirements

The system is designed to be lightweight and accessible, capable of operating on CPU-based systems.

Component	Recommended Specification
Processor	Intel Core i5 or higher
RAM	8 GB or more
Storage	20 GB SSD storage
GPU (Optional)	NVIDIA CUDA-enabled GPU for faster training

The software stack ensures modularity and efficient execution.

Category	Software / Tool Used
<b>Backend Framework</b>	FastAPI (Python 3.10+)
<b>Frontend Stack</b>	HTML5, CSS3, JavaScript
<b>Database</b>	MongoDB
<b>Model Libraries</b>	TensorFlow, PyTorch, Scikit-learn
<b>Audio Processing</b>	Librosa

### Backend and Database Implementation

The backend, built on **FastAPI**, is chosen for its asynchronous processing, speed, and automatic data validation using Pydantic. It handles API requests, model inference, and real-time response generation.

**MongoDB** is utilized as the database, offering a NoSQL document-based structure that is highly flexible and scalable. Key collections store user login credentials (users), emotion predictions (emotion records), and feedback (feedback).

### Functional Flow of the System

The system follows a defined process flow:

14. **Input Collection:** Accepts text, voice (.wav), or handwriting (uploaded image).
15. **Preprocessing:**
  - Text: Tokenization, stopword removal, and sentiment feature extraction.
  - Voice: Extraction of MFCCs, spectral contrast, and chroma features using Librosa.
  - Handwriting: Image resizing, grayscale conversion, and CNN-based feature learning.
16. **Emotion Classification:** Each preprocessed input is passed through its respective deep learning model to predict emotions (e.g., happy, sad, angry, calm).
17. **Fusion and Visualization:** Results are combined using **weighted fusion**, and the final emotion is displayed on the frontend with percentage accuracy and visualization elements.

## Module Implementation Details

### Dataset and Preprocessing Module

This module is foundational for training and testing. It gathers labeled data from various modalities.

- **Text:** Uses datasets like Twitter Emotion Dataset / Kaggle Sentiment140.
- **Voice:** Uses RAVDESS Emotional Speech Dataset.
- **Handwriting:** Uses EMOTIC Handwriting Emotion Dataset.

Preprocessing involves cleaning data, normalization, and splitting into training (80%) and testing (20%) sets.

### Emotion Recognition Modules

Module	Core Algorithm/Model	Purpose & Key Steps
<b>Text (Chat)</b>	Transformer-based sentiment model (e.g., BERT, DistilBERT)	Tokenizes text, removes stopwords, applies the model to predict emotion category (joy, sadness, anger).
<b>Voice (Audio)</b>	CNN-LSTM Model	Extracts MFCC, Chroma, Spectral Contrast features using Librosa and feeds them into the CNN-LSTM model to predict emotion (calm, happy, angry, sad).
<b>Handwriting (Image)</b>	CNN Model	Converts uploaded image to grayscale, resizes, normalizes, and feeds into a CNN model (e.g., ResNet-50) to predict emotional state (stress, calm, neutral) based on stroke features.

### Mood Analyzer and Visualization

The **Mood Analyzer Module** collects emotion data from multiple sessions and calculates a weighted average to produce an overall emotional trend or "mood index".

The **Visualization & Frontend Integration Module** uses HTML, CSS, JavaScript, and libraries like **Chart.js** to display the results

dynamically. Users can view emotional summaries via dynamic graphs, color-coded bars, and charts. The frontend communicates asynchronously with the FastAPI backend using JavaScript's fetch() function.

### Advantages and Discussion

#### Advantages of the Proposed System

ManasAI addresses the limitations of existing unimodal and non-scalable systems by introducing several key advantages:

- **Multimodal Emotion Detection:** Combines handwriting, voice, and text data, leading to higher emotion recognition accuracy and prediction reliability.
- **Real-Time Processing:** The FastAPI backend ensures immediate feedback and smooth communication, essential for continuous monitoring.
- **Cross-Platform Accessibility:** The web-based design allows access on any device with a browser, democratizing mental-health technology.
- **Interactive Visual Output:** Emotions are visualized using charts, graphs, and confidence bars, encouraging users to reflect on their emotional state and promoting mental health awareness, aligning with SDG Goal
- **Lightweight and Deployable:** The system can operate efficiently on CPU-based systems with low latency.
- **User Privacy and Security:** The design focuses on user data privacy and securely storing only necessary metadata.

#### Discussion

The successful implementation of ManasAI validates the effectiveness of multi-modal data fusion combined with a modern, high performance web architecture (FastAPI/MongoDB). The system's architecture is modular and extensible, allowing for future integration of advanced modules like facial emotion detection.

The project provides an efficient and non-judgmental platform that serves as an early-warning mechanism for emotional distress, thereby offering a meaningful step towards emotionally intelligent digital systems that empower human well-being. The use of deep learning models like

CNNs, LSTMs, and NLP transformers ensures high accuracy and robustness across diverse user inputs.

## Conclusion

The ManasAI: Mental Health and Emotion Recognition System has been successfully developed and implemented as an intelligent, multimodal AI platform. By integrating deep learning (CNN, LSTM, Transformer) with a robust FastAPI web architecture and MongoDB, the system offers an efficient, real-time, and user-friendly emotional analysis experience.

ManasAI achieved the following key outcomes:

- Real-time and accurate emotion prediction through multimodal fusion.
- Seamless integration between multiple AI models and the FastAPI/MongoDB stack.
- A clean, responsive, and accessible frontend design.
- Alignment with global wellness objectives, supporting psychological well-being through continuous emotional tracking.

This project successfully bridges the gap between technology and emotional intelligence, offering a significant contribution to the field of AI in healthcare.

## References

1. [1] E. Cambria, S. Poria, R. Bajpai, and B. Schuller, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1515–1521, 2019.
2. [2] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavioral therapy using a conversational agent (Woebot): A randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, e19, 2017.
3. [3] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 2236–2246, 2018.
4. [4] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 216–225, 2014.
5. [5] B. Inkster, P. Sarda, and J. Subramanian, "Digital mental health and chatbots: Therapeutic potential and limitations," *JMIR Mental Health*, vol. 5, no. 1, e19, 2018.
6. [6] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, "Emotion recognition in conversation: Research challenges, datasets, and recent advances," *IEEE Access*, vol. 7, pp. 100943–100953, 2019.
7. [7] J. Calvo, R. Milne, M. Hussain, and H. Christensen, "Natural language processing in mental health applications using non-clinical texts," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 328–341, 2021.
8. [8] R. Taelman, P. Demeester, and F. Ongenae, "AI-powered conversational agents for mental health support: Design, challenges, and evaluation," *IEEE Internet Computing*, vol. 25, no. 6, pp. 45–53, 2021.