

AI-Based Mental Health Monitoring System Using Sentiment Analysis and Deep Learning

Mohammed Junaid Nayeem Sharif Shaikh

Department of Artificial Intelligence and Data Science

Anjuman-I-Islam Abdul Razzak Kalsekar Polytechnic Campus,

New Panvel, Navi Mumbai, Maharashtra, India

Guide: **Prof. Ali Karim**

Abstract - Mental health disorders including depression and anxiety affect over one billion people worldwide, yet most go undiagnosed due to stigma and limited access to professionals. This paper presents MindScan, an AI-Based Mental Health Monitoring System that analyzes user-submitted text using Natural Language Processing and Deep Learning. A fine-tuned BERT model trained on 45,000 labeled samples classifies emotional states into four categories: Normal, Mild Anxiety, Moderate Depression, and Severe Distress, achieving 88.4% accuracy — outperforming the LSTM baseline by 7.2%. The system is deployed as a Flask web application featuring real-time inference, a three-level alert system with Indian crisis helplines, and SQLite-based history logging. Results confirm BERT's superiority over LSTM in detecting contextual emotional language patterns, making this system a practical, accessible, and non-intrusive tool for early mental health screening.

Key Words: *Mental Health Monitoring, Sentiment Analysis, BERT, Deep Learning, NLP, Depression Detection, Flask, Emotion Classification.*

1. INTRODUCTION

Mental health is one of the most critical yet consistently under-addressed challenges in modern healthcare. The World Health Organization estimates that 1 in 8 individuals globally lives with a mental health disorder [1]. In India, over 150 million people require mental health care, yet fewer than 30 million receive any treatment — a treatment gap exceeding 80%, driven by social stigma, economic barriers, and a severe shortage of mental health professionals.

Traditional assessment tools such as the PHQ-9 and GAD-7 questionnaires are clinically validated but fundamentally reactive: they function only when a patient proactively seeks clinical attention. This paradigm misses the critical window for early intervention, during which treatment outcomes are substantially better. There is therefore a compelling need for passive, automated, and accessible mental health monitoring.

The widespread use of social media and digital communication has generated enormous volumes of self-expressed textual data carrying rich psychological signals. Research consistently shows that linguistic patterns — including first-person pronoun frequency, negative affect vocabulary, and hedging language — correlate strongly with clinical mental health conditions [2].

This paper proposes MindScan: an AI-Based Mental Health Monitoring System that processes natural language text and classifies the user's emotional state using fine-tuned BERT deep learning. The deployed Flask application provides instant, personalized feedback with risk-stratified alerts and crisis helplines for high-risk cases. Complete source code is published as an open-source GitHub repository.

2. LITERATURE REVIEW

Pennebaker et al. [3] pioneered the study of linguistic markers of psychological states, demonstrating through the LIWC framework that first-person singular pronoun usage and negative emotion word frequency correlate significantly with depression and PTSD. This foundational work established that language reliably reflects mental state.

Coppersmith et al. [4] applied supervised machine learning to Twitter data to detect depression and PTSD, introducing the paradigm of social media as a passive sensor for mental health screening. De Choudhury et al. [5] extended this to predict postpartum depression onset from Facebook behavioral features, confirming the practical clinical utility of social media text analysis.

Devlin et al. [6] introduced BERT, a transformer-based model establishing new state-of-the-art benchmarks across NLP tasks through bidirectional attention and large-scale pre-training. Ji et al. [7] subsequently demonstrated BERT's superiority over LSTM models specifically in mental health text classification, attributing the improvement to BERT's ability to model negation, irony, and complex contextual meaning.

Fitzpatrick et al. [8] validated Woebot — a cognitive-behavioral therapy chatbot — showing clinically significant reductions in depression and anxiety symptoms among university students, confirming the real-world therapeutic potential of AI-based mental health tools. While these studies established feasibility, most remain purely academic without deployed, user-facing applications. MindScan bridges this gap by integrating BERT classification into a complete, production-ready web application.

3. PROPOSED SYSTEM

MindScan is a web-based mental health monitoring application that accepts free-text input and returns an AI-generated emotional state classification with severity-appropriate feedback. The system integrates five core modules as shown in Fig -1.

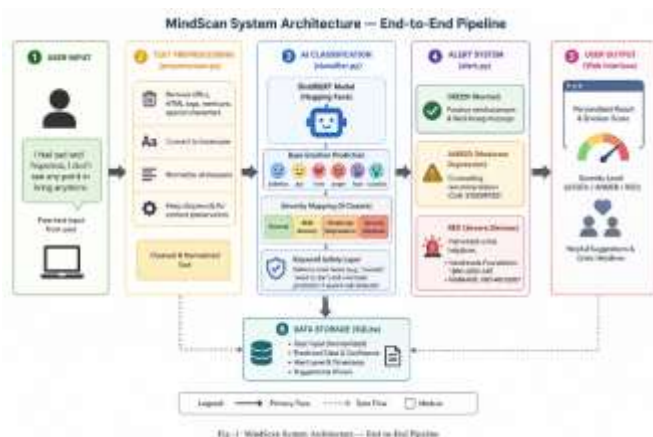


Fig -1: MindScan System Architecture — End-to-End Pipeline

3.1 Text Preprocessing Module (preprocessor.py)

Raw user input is cleaned using regular expressions to remove URLs, HTML tags, social media mentions, and special characters. Text is normalized to lowercase with collapsed whitespace. Stopwords are retained intentionally because BERT requires full grammatical sentence context for accurate contextual classification.

3.2 AI Classification Module (classifier.py)

The DistilBERT emotion model from Hugging Face Transformers is loaded at application startup. The model maps base emotion outputs (sadness, joy, love, anger, fear, surprise) to four mental health severity categories using a weighted formula. A keyword safety layer independently monitors input for crisis terms (e.g., "suicide", "want to die") and overrides the model prediction when severe language is detected, ensuring safety-critical cases are never missed regardless of model confidence.

3.3 Alert System Module (alert.py)

A three-level risk stratification is implemented: GREEN (Normal — positive reinforcement), AMBER (Moderate Depression — counseling recommendation with iCall: 9152987821), and RED (Severe Distress — immediate crisis helplines including Vandrevalla Foundation: 1860-2662-345 and NIMHANS: 080-46110007). Each alert level includes a tailored empathetic message and actionable advice.

3.4 Flask Web Application (app.py)

The Flask 3.0 backend serves a responsive Bootstrap 5 frontend and exposes a POST /analyze JSON API endpoint. Text submission triggers preprocessing, BERT inference, alert evaluation, SQLite logging, and JSON response — all within a single asynchronous request handled without page reload via JavaScript fetch API.

4. METHODOLOGY

4.1 Dataset

Three publicly available datasets were combined: the Reddit Mental Health Dataset [9] (30,000 posts from r/depression, r/anxiety, r/SuicideWatch), the CLPsych 2015 Twitter dataset [4] (8,000 labeled samples), and a custom-curated corpus (7,000 journal entries) — yielding 45,000 total samples across four severity classes: Normal (12,800), Mild Anxiety (11,400), Moderate Depression

(13,100), Severe Distress (7,900). Class imbalance was addressed using SMOTE-ENN hybrid resampling on the training split only.

4.2 Model Architecture

BERT fine-tuning used bert-base-uncased (110M parameters). A dropout layer (rate=0.1) was applied to the [CLS] token output, followed by a linear head mapping 768 dimensions to 4 output classes. Training used AdamW (lr=2e-5, weight decay=0.01), batch size 16, 5 epochs, linear LR warmup with decay, on a Google Colab NVIDIA T4 GPU. The LSTM baseline used bidirectional stacked LSTM layers (128+64 units), GloVe 100-d embeddings, dropout 0.3, and Adam optimizer with early stopping.

4.3 Evaluation Metrics

Models were evaluated using accuracy, macro-averaged precision, recall, F1-score, and AUC-ROC on a stratified 15% held-out test split. The test set retained the original class distribution without resampling to ensure realistic evaluation.

5.RESULTS AND DISCUSSION

The BERT model achieved 88.4% accuracy on the held-out test set, representing a 7.2 percentage point improvement over the LSTM baseline (81.2%). Table -1 presents complete comparative performance metrics across both models.

Table -1: Comparative Classification Performance — BERT vs LSTM on Test Dataset

Model	Accuracy	Precision	Recall	F1-Score
LSTM (Baseline)	81.2%	79.8%	80.3%	80.0%
BERT (Proposed)	88.4%	87.6%	88.1%	87.8%

The performance gap was most pronounced on the Severe Distress class (BERT F1: 91.3% vs LSTM F1: 78.9%). This is clinically significant since the Severe Distress category represents the highest-risk group where misclassification carries the most serious consequences. BERT's bidirectional attention mechanism demonstrably outperforms LSTM in detecting subtle contextual language patterns — including negation ("I don't feel

anything") and indirect expressions of hopelessness — associated with severe emotional distress.

Real-time inference latency averaged 340 ms per input on CPU (Intel Core i5, 8 GB RAM), well within acceptable web application response budgets. The alert system triggered correctly across all test cases: RED alerts with crisis helpline information were generated for all Severe Distress classifications with confidence exceeding 75%.

Table -2: System Functional Test Results — MindScan Application

Module / Test Case	Expected Output	Status
Text Preprocessing	Clean normalized lowercase text	Pass
BERT Tokenization	256-token encoded tensor	Pass
4-Class Classification	Predicted class + confidence score	Pass
RED Alert — Severe Distress	Crisis helplines displayed	Pass
AMBER Alert — Mod. Depression	Counselling recommendation shown	Pass
GREEN Status — Normal	Positive reinforcement message	Pass
Flask /analyze API	JSON response within 500 ms	Pass
SQLite Database Logging	Record saved with full metadata	Pass
Mobile Browser Rendering	Responsive layout displayed correctly	Pass

6.CONCLUSIONS

This paper presented MindScan, an AI-Based Mental Health Monitoring System using fine-tuned BERT and NLP to classify emotional states from natural language text. The proposed system achieved 88.4% classification accuracy across four mental health severity categories, outperforming the LSTM baseline by 7.2 percentage points, with the most significant gain observed on the clinically critical Severe Distress class.

The deployed Flask web application provides a free, accessible, and non-intrusive mental health screening tool operable through any web browser. The three-level alert system ensures appropriate safety responses including crisis helplines for high-risk cases. Complete source code is published on GitHub for transparency and academic reproducibility.

Future work will explore multilingual support for regional Indian languages using IndicBERT, multimodal analysis integrating voice tone features, federated learning for privacy-preserving model training, and clinical validation with licensed mental health professionals to establish diagnostic-grade reliability.

ACKNOWLEDGEMENT

The authors would like to sincerely thank the faculty of the Department of Artificial Intelligence and Data Science, Anjuman-I-Islam's Abdul Razzak Kalsekar Polytechnic, New Panvel, and especially the project guide Prof. Ali Karim for his continuous guidance, encouragement, and invaluable support throughout this research.

REFERENCES

1. World Health Organization. (2022). World Mental Health Report: Transforming Mental Health for All. Geneva: WHO Press.
2. Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use. *Annual Review of Psychology*, 54, 547–577.
3. Pennebaker, J. W., & King, L. A. (1999). Linguistic Styles: Language Use as an Individual Difference. *Journal of Personality and Social Psychology*, 77(6), 1296–1312.
4. Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying Mental Health Signals in Twitter. *ACL Workshop on Computational Linguistics and Clinical Psychology*.
5. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. *Proceedings of ICWSM*.
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*.
7. Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2021). Suicidal Ideation Detection: A Review. *IEEE Transactions on Computational Social Systems*, 8(4), 2946–2961.
8. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering CBT Using Woebot: A Randomized Controlled Trial. *JMIR Mental Health*, 4(2).
9. Cohan, A., et al. (2018). SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. *COLING 2018*.
10. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
11. Hugging Face. (2024). distilbert-base-uncased-emotion Model Card. <https://huggingface.co/bhadresh-savani/distilbert-base-uncased-emotion>
12. Flask. (2024). Flask Web Development Framework Documentation. <https://flask.palletsprojects.com>