

Emo Sensei - Mood and Stress Level Detection Based on Artificial Intelligence Using Voice, Text, And Video Analytics

Pawan Goyal

Bachelor of Technology

Vivekananda Global University

Harshit Kumar

Bachelor of Technology

Vivekananda Global University

Thota Vijaya

Bachelor of Technology

Vivekananda Global University

Ishani Jain

Bachelor of Business Administration

Vivekananda Global University

Faculty of Computer Science and Engineering

Superintendent: **Dr. Gyan Prabhakar** | Academic Year: After 2025-2026

Abstract

With the increasing levels of stress and psychological distress, the need for effective and easy-to-use monitoring systems is already evident. This article describes EmoSense AI, a multi-faceted system that analyzes voice and text data to detect mood and stress levels. Unlike traditional approaches that rely on physical sensors, EmoSense AI uses natural, everyday conversations. This makes it realistic, scalable, and suitable for real-world use.

The system processes acoustic features extracted from speech signals, such as timbre, power, and multi-frequency cepstrum coefficients (MFCCs). In addition, it processes semantic and emotion-related features extracted from text using natural language processing (NLP) techniques. Emotion classification is performed using a deep learning framework. A long-term memory network (LSTM) processes the text input, and a hybrid convolutional neural network model (LSTM) processes the speech. In the next step, the results of both methods are combined using a “delayed fusion” technique to improve the overall reliability and accuracy of the recognition process. [2], [3]

The results showed that the combined use of voice and text helped to better understand emotional states than voice or text-based methods alone. From student mental health monitoring and workplace wellness programs to intelligent conversational agents, EmoSense AI has a wide range of potential applications. In addition, it has the potential to become a proactive tool for early diagnosis of stress.

Key Insight

By moving from single-channel to dual-channel analysis, EmoSense AI can now understand emotional states more comprehensively, accurately, and in near real-time. As a result, it is a much more reliable solution than traditional single-channel approaches.



Lack of visibility in mental health care

About spirituality

Mental illness is a serious and chronic problem because it is often invisible and private. Physical illnesses are often identified through visible symptoms or diagnostic tests, but psychological conditions such as stress, anxiety, and depression are often not diagnosed until they reach a more severe stage. This chronic problem is called the mental health visibility gap. [11]

This gap is growing rapidly in today's digital world. With the reliance on chat, text messaging, email, and virtual media, it has become easier than ever for people to hide their problems. Because traditional behavioral indicators such as facial expressions, gestures, and tone of voice are often absent in digital interactions, it becomes difficult for doctors and mental health professionals to objectively and consistently assess a person's emotional state.

These challenges are compounded by a major problem called self-reporting. Due to social bias, fear of criticism from others, or a lack of self-awareness, people often underestimate or misrepresent their true feelings. Vague responses such as "I'm fine" often mask serious internal emotional distress, leading to misdiagnosis and missed opportunities for early treatment.

Delays in treatment are also a concern. In many cases, people only seek professional help after symptoms have significantly worsened; at this stage, preventive measures are ineffective. Therefore, the development of reliable methods of early diagnosis is not only beneficial, but also essential to prevent long-term psychological and physical damage.

Research has consistently shown that many people with mental illnesses do not seek professional help. It is estimated that approximately 60% of people avoid professional help, a situation that is exacerbated by stigma, lack of information, and barriers to accessing care. [12]

Overall, there is a strong indication of the need for automated, targeted, and continuous monitoring systems. Artificial intelligence-driven approaches, particularly those based on behavioral cues found in everyday conversation and writing, show a promising path for the future. [7]

Dual channel detection and real-time evaluation

Emocence listens to speech in a single-channel intervention framework, and analyzes its acoustics and capabilities to provide more accurate and concise intervention. The acoustic model captures physiological signals in the voice, such as pain, vibration, and shock, which are associated with emotional arousal and stress. Similarly, the linguistic model uses NLP techniques to interpret cognitive concepts by analyzing the conversation, basic syntax, and linguistic structure to

determine the mood of the conversation . [10] These two channels are combined in a hybrid CNN-LSTM fusion model, which requires simultaneous local dynamic learning. By combining complementary signals from speech and writing , this fusion technique achieves speech understanding. The system fully preserves the information in this conversation and ensures that all important information is processed without interference.

About the EmoSense Mobile App

Emo Sense AI provides simple mental health monitoring and support by seamlessly integrating into your daily life through various mobile services. The app is designed to fit seamlessly into your daily life, serving as both a passive monitoring tool and an active health platform. It collects data in the background and provides a well-organized health interface that users can directly engage with.

One of its key features is the “Daily Voicemail,” where users can record short voice messages. The recorded voices are automatically analyzed to identify emotional states and signs of stress . This simple and user-friendly model increases the practical utility and sustainability of the platform. It also includes a convenient mood dashboard, which shows mood changes on a daily, weekly, and monthly basis. This gives both users and caregivers a clear understanding of long-term behavioral patterns.

This system is designed with privacy and legal use in mind. Data processing within the device ensures the security of personal data and increases user confidence in the use of mental health support tools.

In short, these systems effectively transform mental health research goals into practical and effective tools for everyday use, taking advantage of the widespread availability, ease of access, and monitoring capabilities of mobile technology.

System Architecture And Technology Stack

This system is built on a robust and comprehensive architecture to ensure unparalleled implementation and reliability.

About Mobile Frontend

- Response core (multi-dimensional development).
- After redox (state management).

Backend AI Center (Python-based).

- TensorFlow for deep learning models [9].
- Le Pros for audio signal processing [4].
- Transformer-based models for text analysis [5], [6].

After the main work

Real-time processing of data will be performed.

After properly extracting the features.

- Efficient sampling performance

By securely managing users' confidential information, we create an environment that is conducive to continuous mental health monitoring.

Data Flow And Processing Pipeline

This system follows a structured information flow algorithm, which has four main stages: input, processing, analysis, and feedback.

Entry level

Users provide information by recording voice in the app.

It serves as the main source of information for analyzing emotions and stress.

Processing method

- The recorded sound is improved by removing noise and unnecessary signals.

The processed data is sent securely for further calculations.

Analysis phase

The artificial intelligence model includes many features such as phonetic patterns and linguistic patterns.

These features are used to identify the user's psychological state and stress level.

Verb tense: hola

- Draw conclusions based on system analysis.

Users can get real ideas and suggestions to solve the problem.

This method allows for effective, continuous, and real-time mood monitoring.

Scientific Stress Level Classification

This system uses a quantitative scoring system to classify stress levels and objectively assess psychological state. [15]

Stress levels are divided into five categories based on the range of scores. [13]

0-14 (Normal): Indicates a stable and healthy mental state.

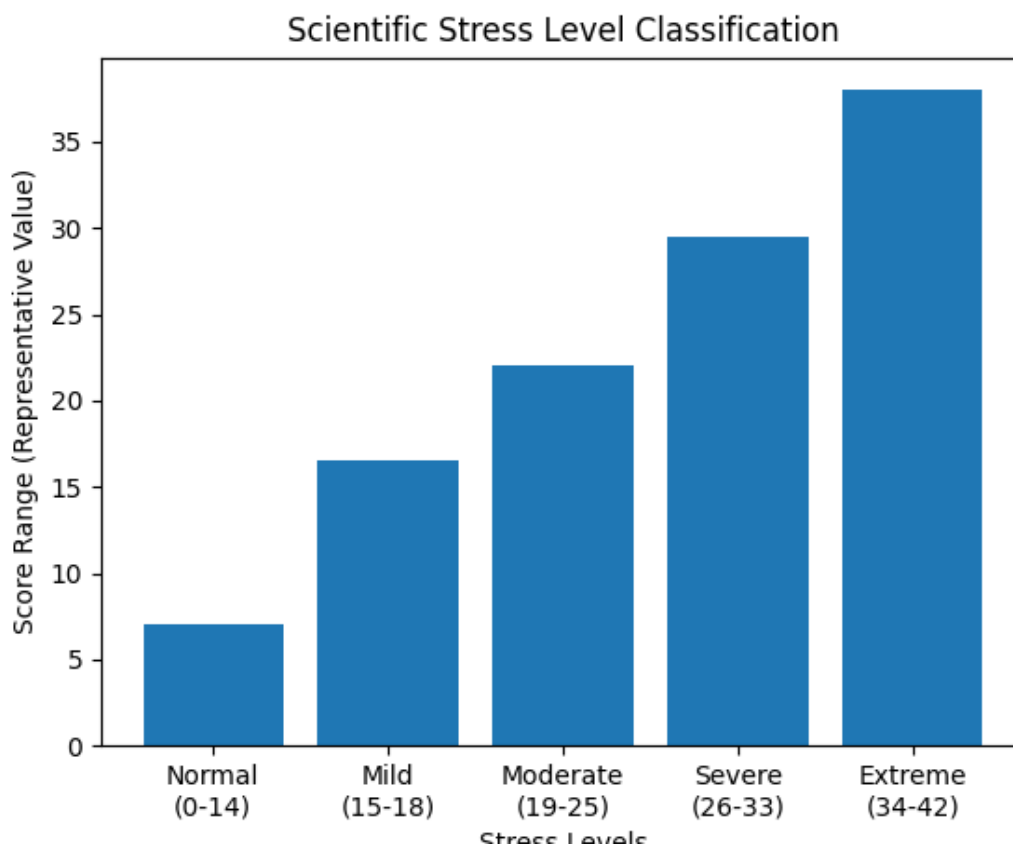
- 15-18 (moderate): Indicates moderate stress with manageable consequences.

- 19-25 (moderate): Signs of obvious stress are evident and attention is needed.

©26-33 (Severe): Symptoms of high blood pressure and may have adverse effects.

34-42 (Very Severe): Indicates severe stress, which requires immediate intervention.

This classification is based on the analysis of a combination of phonological and linguistic features. This structural classification provides a clear understanding of the user's situation, helps to provide personalized feedback and recommendations, and helps to identify high-risk events in a timely manner. This model ensures consistent and standardized stress assessments across different user groups.



Empowering Clinicians And Clinical Decision Support

Emo Sense's AI aims to not only help individuals, but also empower mental health professionals through its unique clinical dashboard. This dashboard allows clinicians to continuously monitor patients' health, even during formal treatment. By combining patient-provided data with objective and data-driven insights, it helps bridge the long-standing gap between patient information and clinical observation.

Key Functionalities

Analysis of objective data to confirm and validate the symptoms described by patients.

It is useful to monitor the patient's progress between treatments.

This provides early warning of potential risks such as severe stress or a relapse of symptoms.

- It allows the effects of treatment to be monitored over time, which helps make more accurate medical decisions.

Impact

Emo Sense AI will seamlessly integrate AI-powered insights into clinical workflows to improve decision-making, facilitate early intervention, and make personalized treatment plans more effective. It will create a model of mental health practice that is predictive, evidence-based, and ultimately more effective.

Market Growth And Industry Potential

The mental health technology sector is experiencing significant growth. The rapid growth of e-commerce, increasing public awareness, the rapid spread of digital devices, and the availability of accessible mental health solutions are fueling demand for them worldwide. According to recent market research estimates, the mental health app market is expected to reach an estimated \$17.5 billion by 2030, indicating continued and rapid growth in demand across diverse populations and regions. [14]

Title

This market is growing steadily and is estimated to grow from \$5.2 billion in 2023 to \$17.5 billion by 2030.

The following factors are responsible for this significant development:

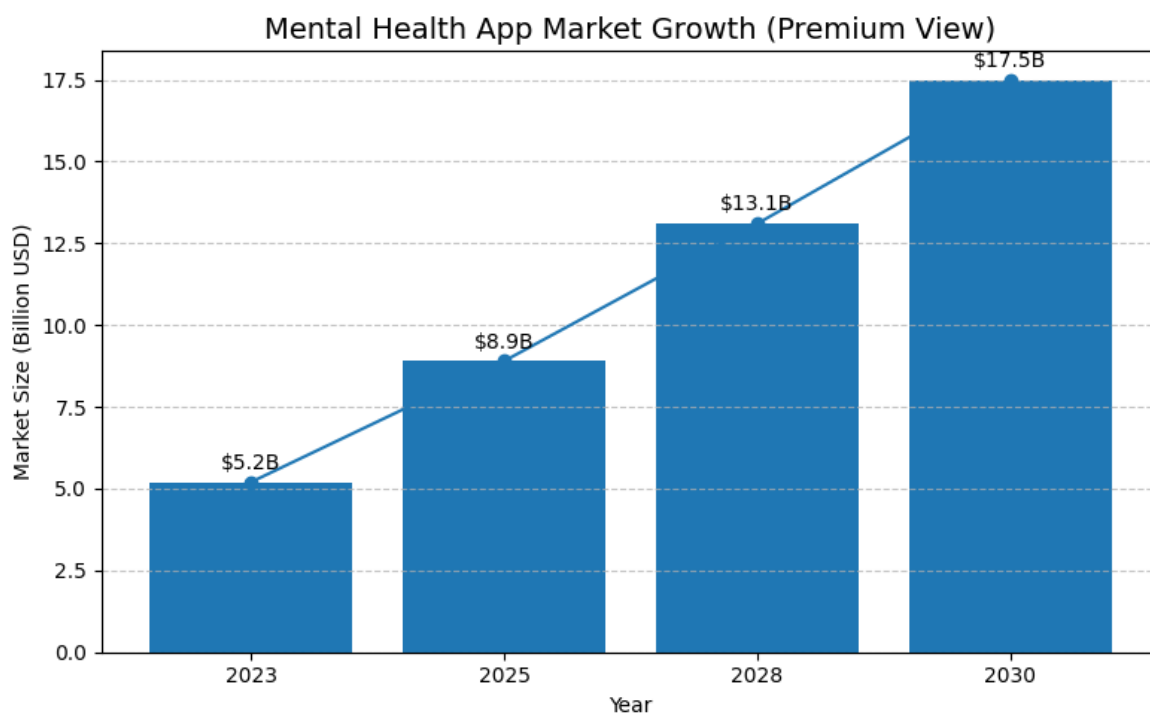
Raising awareness about mental health issues

The use of smartphones and mobile applications is increasing.

There is a growing demand for solutions that allow for the delivery of telemedicine and digital healthcare.

Medium-term forecast: \$8.9 billion by 2025 and \$13.1 billion by 2028.

M-Sense Artificial Intelligence is a huge opportunity. The demand for scalable, scalable, and intelligent mental health solutions is growing rapidly worldwide. M-Sense Artificial Intelligence is uniquely positioned to meet this urgent need.



Comparison: Existing AI Vs. EmoSense AI

Currently, most AI-based mental health tools work in a similar way, relying on both plain text (written words) and audio data. While each approach has provided valuable information, none has fully captured the depth and complexity of human emotions. Text-based systems often fail to recognize the subtleties of speech and nonverbal cues, and audio-based systems often fail to fully understand the subtleties and meanings hidden within language. To overcome this fundamental shortcoming, we built Emusense AI into a dual-channel architecture, combining it with a single, integrated approach to speech and language analysis. [8]

Comparison of Existing AI Systems vs. EmoSense AI

Feature	Existing AI Systems	EmoSense AI
Data Input	Single-channel (Text or Voice)	Dual-channel (Voice + Text)
Analysis Type	Limited (only sentiment or tone)	Combined acoustic + linguistic analysis
Emotional Understanding	Partial	Holistic (physiological + cognitive)
Accuracy	Lower (higher false positives)	Higher (cross-validation of data)
Context Awareness	Limited or missing	Strong contextual + behavioral understanding
Detection Speed	Delayed or reactive	Real-time and proactive
Reliability	Moderate	High

References

- [1] S. Boria, E. Cambria, R. Bajpai, A. Hussain, "Review of sentiment analysis: From single-method analysis to multi-method fusion", *Information Fusion*, vol. 1, pp. 101-37, pp. 98-125, 2017.
- [2] One.
- [3] S. Hoekreiter and J. Schmittuber, "Long-term memory", *Neural Computation*, Volume 9, Issue 8, pp. 1735-1780, 1997.
- [4] P. McAfee et al., "Librosa : Speech and Music Signal Analysis with Python," *Proceedings of the 14th Python in Science Conference*, Austin, Texas, USA, 2015, pp. 18–25.
- [5] A. Vaswani et al., "Attention is everything", *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 1, p. 101. See also *ba on 30*, pp. 5998-6008.
- [6] J. Devlin, M. W. K. Ba. Chang, K. Li, K. Dotanova, "BART: Pre-training a Deep Bidirectional Transformer for Language Understanding," *Proc. NAACL-HLT*, Minneapolis, Minnesota, June 2019, pp. 4171–4186.
- [7] R. Koe et al., "Emotion Recognition in Human-Computer Interaction", *IEEE Journal of Signal Processing*, Volume 18, Issue 1, pp. 32-80, January 2001.
- [8] Z. Zhang, "Towards robust multi-modal sentiment analysis", *Proceedings of the ACM ICMI*, Boulder, Colorado, October 2018, pp. 347-355.
- [9] M. Abadi et al., "TensorFlow: A Framework for Large-Scale Machine Learning", *Proceedings of the 12th USNIX OSDI Conference*, Savannah, Georgia, November 2016, pp. 265-283.
- [10] KR Scherer, "Verbal Emotion Expression: A Review and Model for Future Research", *Psychological Bulletin*, Volume 99, Issue 2, pp. 143-165, 1986.
- [11] World Health Organization, "Depression and other common mental disorders: global health estimates", WHO, Geneva, Technical report. WHO/MSD/MER/2017.2, 2017. [Online]. Available at: <https://www.who.int/publications/i/item/depression-global-health-estimates>
- [12] R. C. Kessler et al., "Lifetime prevalence of DSM-IV disorders and distribution by age of onset in a replication of the Nationwide Comorbidity Survey," *Archives of General Psychiatry*, vol. 1, p. 101. Vol. 62, no. 1, issue 6, pp. 593-602, June 2005.
- [13] P. Ekman, "The Debate of Basic Emotions", *Perception and Emotion*, Volume 6, Issue 3-4, pp. 169-200, 1992.
- [14] Grandview Research, "Mental Health Apps Market Size, Share and Trends Analysis Report", *Market Research Report GVR-4-68038-955-9*, 2023. [Online]. Available at: <https://www.grandviewresearch.com/industry-analysis/mental-health-apps-market>
- [15] M. Loibond and P. F. Loibond, "The structure of negative emotional states: a comparison of the Depression-Anxiety-Stress Scale (DASS), the Beck Depression-Anxiety Scale, and the Beck Anxiety Scale", *Behavioral Research and Therapy*, 33, No. 1, pp. 335-343, 1995.