A Survey on Cross-Platform Depression Detection: Combining Text, Audio, Images to Understand Emotions over Time

Anuradha Muttemwar ; Shruti Shivankar ; Lisha Gabhane
Dept of MCA, GHRCEM, Nagpur, Maharashtra, India

Abstract

Depression is a serious mental health condition that generallysee a unnoticed or untreated because there is not to much awareness about it. Most people don’t really understand what it is, and it’s not always diagnosed using proper or traditional methods.Traditional approaches, such as personal evaluations and doctor consultations, is typically subjective and influenced by external factors like mood, surroundings, or social pressure. This survey attempts to explore a cross-platform system for detecting depression by analyzing text, audio, and images. By conducting systematic and well-structured survey, we hope to understand public perceptions of AI-based mental health monitoring anddetermine how ready theyare to accept such te chnology.The survey highlights the importance like the accuracy of text sentiment analysis, speech emotion recognition, and facial expression detectionthrough images to identify patterns related to depression. It also looks user about regarding privacy, trust, and the ethical use of AI-based detection systems. By gathering insights from various participants, this study offers a clear understanding of public attitudes toward multimodal depression detection method. The findings of this survey may provide to future advancements in AI-based mental health monitoring, leading to a more precise and consistent understanding of emotional well-being.

Keywords:
Mental Health, Depression, Multimodal detection, Sentiment Analysis

1. Introduction

Depression is a mental health condition that globally affects millions of people and that manifests through random changes in behavior, speech, and appearance. In the old days, the traditional diagnostic methods were dependent on self-reports, and doctor's assessments which are influenced by different external factors, that might be subjective [1]. In this research, we introduce a cross-platform approach where we combine data via text, images and audio.

By analyzing this modularity, we can try to study a complete view of emotional states over time. Through textual data, we can analyze the patterns, by audio vocal tone and pitch captured and through images we get insights into facial expressions or body language [2,4]. This multi-modal approach seeks to improve depression detection and track symptom progression this will provide a more objective and continuous system for tracking emotional well-being. This work presents a new approach to depression detection by integrating multi-modal data from many platforms, such as audio, image, and text sources. By studying these many emotional expressions, we hope to create a more robust and dynamic framework for understanding emotions over time [5].

A more precise and comprehensive knowledge of depression symptoms ismade possible by the integration of speech sentiment analysis, facial expression recognition, and natural language processing (NLP). This study investigates the potential of cross-platform data to enhance the precision of depression diagnosis and continually track emotional shifts, providing a viable means of early identification and tailored treatment [8].

DOI: https://doi.org/10.5281/zenodo.15462055

Its ability to track depression over time is one of the key advantages of this cross-platform strategy. Depressive disorders are, however, often undiagnosed or untreated because of denial by certain patients and a lack of adequate awareness of the condition in much of the world. Many studies agree that social media platforms, where people can openly express their views and emotions, can be important for tracking health issues and trends [2]. Still, as people do not put everything on the internet, their engagement activities like what they scroll through or social interaction and their browsing like what they view on their phones can also help in detection while ensuring their privacy is protected [13]. Multimodal detection systems can continually examine emotional patterns across many media, providing a deeper, more complex profile of a person's mental health than traditional approaches that might rely on a snapshot assessment during a single contact. Early detection, individualized therapeutic approaches, and more precise treatment progress tracking are made possible by this [8].

2. Research Approach

2. 1literature Review

Artificial Integration is now being involved in monitoring mental health as well. Previously, detecting depression was dependent on physicians and whether someone's mood and social stress influenced what they disclosed or not. Presently, AI can identify depression based on analyzing text, audio, and visuals without bias [1, 3].

1) Role of AI in Mental Health Detection

Artificial intelligence-based tools such as text sentiment analysis, audio recognition, and facial expression detection are evaluated for identifying depression. Challenges like privacy of data, accuracy, and ethics are still a concern [9].

2) Do People Trust AI?

The survey adopts a scientific process:

- Survey Creation: Crafting applicable questions
- Distribution: Dissemination through Google Forms, social media, and face-to-face interviews
- Data Collection: Accumulating replies
- Analysis: Determining patterns and behaviour trends
- Data Collection Methods

Replies will be gathered through Google Forms, social media surveys, and face-to-face interviews for a diverse array of participants and true-world feedback.

Some individuals prefer AI-based monitoring because it's accessible 24/7, but others are still wary about privacy and trust. They're afraid of data security threats and the dependability of AI forecasts [6,11].

Table.1. Summary of Previous Surveys on AI-based tools for Detecting Depression

Source: Adapted from [2], [4], [7], [11], [13]

Focus of Survey	**Year**	**Observations or Results**	**Challenges**
Sentiment analysis	2023	Detecting emotions that are positive/negative.	Analyze emotion [4]
Text Recognition	2021	It can detect social media through analysis of text-based activity.	Availability of Sufficient data [11]
Image Recognition	2022	AI detects depression through image.	Balance different type of data [7].
Audio Recognition	2023	AI detect emotions over voice	Needs clear voice [2]
Public Perception on AI	2020	People are concerned about AI	Concerns about security and privacy

DOI: https://doi.org/10.5281/zenodo.15462055

2.2 Survey Design & Methodology

The main aim of this survey is to gauge the public attitude towards AI-based depression detection, whether people believe in such systems or are apprehensive about privacy and ethics. To do this, a structured survey has been developed to measure public awareness, level of acceptance, and ethical aspects of AI-driven mental health surveillance [13].

1) Target Audience

The survey comprises members from different age groups and geographies to get a rich array of opinions. The sample size is intentionally selected to offer significant data analysis without allowing biased assumptions.

2) Survey Design

The survey is broken down into three main categories:

- Awareness: Are individuals aware of AI-enabled mental health detection?
- Trust & Acceptance: Do people find AI predictions trustworthy?
- Ethical Concerns: Is there any fear of invasion of privacy, data safety, or misuse?

3) Survey Process

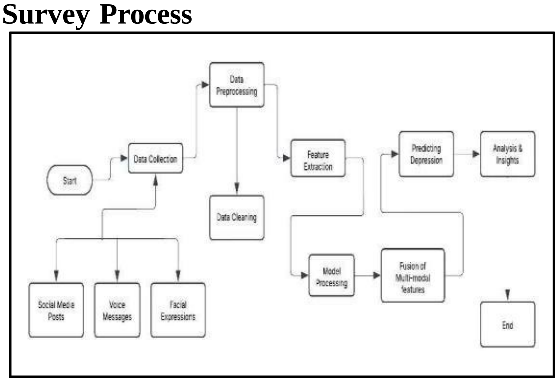

Figure 1: Survey Process for Cross-Platform Depression Detection

4) Sample Size Justification

The selected sample size is made with the intention to provide statistical validity and a lucid understanding of public attitudes toward AI-based mental health monitoring. A small sample can produce biased results, while an overly large sample can prove to be infeasible. Hence, the right balance is struck to ensure reliability and feasibility.

This survey hopes to generate meaningful information towards future development in AI-based depression detection in line with tackling issues of concern in public perceptions around trust, acceptance, and ethics.

3. Data Collection & Analysis

3.1 collecting Data From Cross-Platform Approach

1) **Twitter API:** To gather textual data, we utilized the Tweepy library to gather tweets in real time. This allowed us to monitor certain keywords or hashtags, such as "depression" or "mental health," and retrieve publicly posted tweets [10].
2) **Instagram API:** Instagram data, particularly public posts or comments, can be accessed using APIs or web scraping services. Because Instagram is heavily pictorial, even image captions, hashtags, and comments can be included in sentiment analysis and processed for sentiment by Natural Language Processing (NLP) techniques [5].

3) **Reddit API:** The PRAW (Python Reddit API Wrapper) library was used to collect comments from subreddits covering the scope of the study, such as those focused on mental health. Reddit provides rich textual data that enables in-depth sentiment analysis and discussions on various topics [11].

4) **YouTube API:** YouTube comments were gathered through YouTube's API. In addition to text analysis, sentiment analysis may also be applied to video titles and descriptions. Commentsentiment can be used to measure the emotional tone of publicopinion on different videos, providing insights into emotions or feelings linked to particular content or trends [9].

3.2 Survey Results and Analysis

This section presents the results of the survey carried out to evaluate public attitudes towards

DOI: https://doi.org/10.5281/zenodo.15462055

AI-based mental health surveillance. The survey findings reveal the levels of awareness, confidence in AI-powered depression diagnosis, and privacy and ethics concerns. The results are shown using visual aids like charts and tables for easier understanding.

1) Public Awareness of AI in Mental Health

One of the main goals of this survey was to identify to what extent AI-based mental monitoring is known by the public. A large group of respondents displayed a general notion of AI functionality in mental monitoring, but some were not cognizant of its particular mechanisms in detecting depression.

Public Awareness and Acceptance of AI in Mental Health (2025)

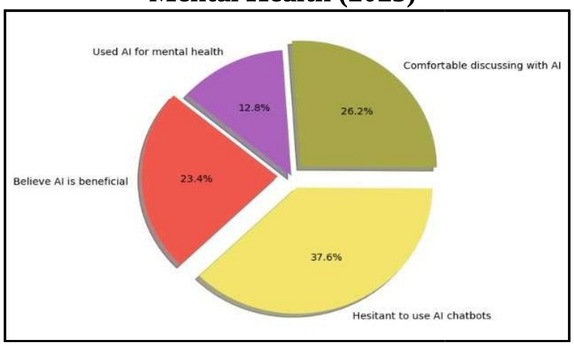

Figure 2: Pie chart illustrating the prevalence of awareness levels among participants.

The information shows that although a significant percentage of the respondents are aware of AI's contribution to mental health, a significant majority is unaware [10].

2) Trust Levels in AI-Based Mental Health Assessment

Confidence in AI-based mental health tracking is a pivotal factor determining its uptake. The questionnaire respondents were posed the question of whether they believed AI can evaluate text sentiment, speech emotion, and facial expressions to identify depression accurately [11].

3) Concerns on Privacy, Accuracy, and Ethical Risks

Participants voiced many concerns about AI depression detection systems. The most common concerns raised were data privacy, the possibility of AI models being biased, and the ethical use of automated systems in mental health assessments [3, 10].

4) AI Acceptance Across Different Age Groups

Acceptance of AI depression detection systems differed across the age groups. Younger age respondents (18-30 years) had greater acceptance levels, while those in the older age group (50+ years) were less accepting.

3.3 Sentiment Analysis for Cross-Platform Approach

1) This study investigates the application of social media for early depression detection through both text-based and cross-platform sentiment analysis. The study addresses several key questions, are:

a. What are the issues in applying text-based approaches to identify depression on social media?
b. Whatisthe most effective text-based methods for detecting early depression? [11]

The Google Cloud Natural Language API was utilized for sentiment analysis on text data from various platforms. To further analyze, cross-platform depression detection techniques can be applied to evaluate sentiment from visual and audio content as well.

2) For images, sentiment can be processed using image recognition and classification algorithms such as CNNs (Convolutional Neural Networks) to identify facial expressions, objects, and scenes that convey mood, which can understand the emotional content of visual material (e.g., facial expressions or objects within the image). Also, color-based sentiment recognition is important in knowing the emotions of that particular person [5,6].

3) for audio information (if it exists, like on podcasts or YouTube), speech-to-text models

DOI: https://doi.org/10.5281/zenodo.15462055

can first translate the audio into text, which can be examined for sentiment via text-based NLP methods [2].

3.4 Features Extraction

1) **Text aspects:** Take the text data and extract lexical (word frequency, sentiment ratings, etc.), syntactic (sentence length, complexity), and semantic (emotional content, cognitive distortions) aspects [10].

2) **Audio Features:** Prosodic characteristics such as pitch fluctuation, speech tempo, vocal intensity, and pauses can be extracted from audio data. Emotional emotions such as despair or apathy are reflected in these traits [2].

3) **Image Features:** Take visual cues from facial expressions, including the facial action units (AUs) that represent happiness, rage, or despair. Emotional states are also deciphered through body language and head posture studies [5].

3.5 Discussion

Public awareness of AI in monitoring mental health is moderate, with a requirement for enhanced education about its advantages and limitations, according to the survey findings [10,11]. Trust in AI-based diagnosis is still low since most respondents are not keen on using AI-based diagnosis. Privacy, accuracy, and ethical concerns are chief obstacles to the full implementation [3,7,10]. Younger respondents are more receptive to AI-driven depression detection, while the elderly are resistant because of doubts regarding the involvement of technology in mental illness [11]. These results highlight the imperative of understandable AI systems with explicit ethical frameworks and human supervision to increase public acceptance and trust [6,12]. Overcoming privacy issues and enhancing AI accuracy using multimodal analysis could lead to increased use of AI-driven mental illness monitoring in the future.

The success of identifying multimodal depression depends on the fusion and processing of different data sources. The combination of text, images, and audio improves the reliability of predictions by addressing the limitations of the unimodal approach [4,8,9]. Sentiment analysis using only text can identify negative emotions but lacks context without facial images or voice data. Facial expression detection alone struggles to capture the full range of depressive mood symptoms, especially those expressed through tone and speech rhythm [5,6].

Feature fusion, or integrating text, audio, and image data, is a key challenge in this approach. They use deep learning techniques like Convolutional Neural Networks (CNN) for images, Recurrent Neural Networks (RNN) for text, and spectrogram-based models for audio classification as the older machine learning models frequently struggle to handle complex multimodal data, these models provide predictions more accurately [4,5,8].

In the future, the focus of this research is to improve the way of combining different types of data, increasing data variety and making sure AI is fair and follows ethical guidelines. Teamwork between computer scientists, psychologists and doctors can improve the accuracy and importance of depression detection models [12]. By solving these challenges, cross-platform depression detection can become a helpful tool for monitoring mental health and providing early support.

4. Conclusion

In early mental health evaluation techniques text, image, and audio multi-modal depression detection is a practical approach. By integrating several data sources, accuracy prediction is increased however, issues including imbalance, data privacy, and feature fusion complexity need to be resolved [4,12]. The diverse collection of datasets trustworthy AI frameworks and moral AI standards are necessary for responsible implementation.

DOI: https://doi.org/10.5281/zenodo.15462055

The main goals of future studies are to enhance fusion methods, explainability, and real-time monitoring. If these techniques are improved, this technology can serve as a valuable tool for mental health support and providing early intervention. It can also reduce disorders like depression and prevent further complications.

References

[1] Bin Tahir, W., Khalid, S., Almutairi, S., Abohashrh, M., Memon, S. A., & Khan, J. (2025). Depression detection in social media: A comprehensive review of machine learning and deep learning techniques. IEEE Access, 13, 12789–12818. https://doi.org/10.1109/ACCESS.2025.3530862

[2] Mamidisetti, S., & Reddy, A. M. (2023). A stacking-based ensemble framework for automatic depression detection using audio signals. International Journal of Advanced Computer Science and Applications, 14(7), 603–612.https://doi.org/10.14569/IJACSA.2023.0140767

[3] Martinez, E., Cuadrado, J., Peña, D., Martinez-Santos, J. C., & Puertas, E. (2023). Automated depression detection in text data: Leveraging lexical features, phone's themes embedding, and RoBERTa transformer model. In IberLEF (Working Notes). CEUR Workshop Proceedings.

[4] Das, R., & Singh, T. D. (2023). Multimodal sentiment analysis: A survey of methods, trends, and challenges. ACM Computing Surveys, 55(13s), Article 270, 1–38.https://doi.org/10.1145/3586075

[5] Mamidisetti, S., & Reddy, A. M. (2022). Multimodal depression detection using audio, visual, and textual cues: A survey. NeuroQuantology, 20(4), 325–336. https://doi.org/10.14704/nq.2022.20.4.NQ22127

[6] Clancey W. J. and Rennels G. R, "Strategic Explanations in Consultation", The International Journal of Man-Machine Studies, Vol. 20, No. 1, pp. 3–19, 1983.

[7] Ignatiev, N., Smirnov, I., & Stankevich, M. (2022). Predicting depression with text, image, and profile data from social media. In Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2022) (pp. 753–760).https://doi.org/10.5220/0010986100003114

[8] Thati, R. P., Dhadwal, A. S., Kumar, P., & Sainaba, P. (2022). A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. Multimedia Tools and Applications, 82, 4787–4820.https://doi.org/10.1007/s11042-022-12315-2

[9] Salas-Zárate, R., Alor-Hernández, G., Salas-Zárate, M. d. P., Paredes-Valverde, M. A., Bustos-López, M., & Sánchez-Cervantes, J. L. (2022). Detecting depression signs on social media: A systematic literature review. Healthcare, 10(2), 291.https://doi.org/10.3390/healthcare10020291

[10] Yoon, J., Kang, C., Kim, S., & Han, J. (2022). D-vlog: Multimodal vlog dataset for depression detection. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12932–12940.https://doi.org/10.1609/aaai.v36i11.21483

[11] William, D., & Suhartono, D. (2021). Text-based depression detection on social media posts: A systematic literature review. Procedia Computer Science, 179, 582–

DOI: https://doi.org/10.5281/zenodo.15462055

589.https://doi.org/10.1016/j.procs.2021.01.043

[12] Ye, J., Yu, Y., Wang, Q., Li, W., Liang, H., Zheng, Y., & Fu, G. (2021). Multi-modal depression detection based on emotional audio and evaluation text. Journal of Affective Disorders, 295, 904–913. https://doi.org/10.1016/j.jad.2021.08.090

[13] Martínez- Castaño, R., Pichel, J. C., & Losada, D. E. (2020). A big data platform for real-time analysis of signs of depression in social media. International Journal of Environmental Research and Public Health, 17(13), 4752. https://doi.org/10.3390/ijerph171347