

Early Depression Detection using AI: A Web Based Psychiatrist-Patient Platform

Shruti Shelke¹, Lavanya Gaikwad², Siddhi Unawane³, Om Sonawane⁴,
Miss. Bharti Ahuja⁵

Guru Gobind Singh College of Engineering and Research Center, Nashik

Abstract

The proposed system is a web portal designed to enhance mental health management by connecting psychiatrists and patients. It offers secure login and personalized dashboards for patients, featuring therapeutic videos on themes like depression and happiness. Using machine learning algorithms, particularly the Random Forest algorithm, the system tracks patient activity, such as video views and interactions, to detect potential mental health concerns like depression. The Random Forest model is employed for its high accuracy and robustness in classification, enabling reliable detection of mental health issues. Alerts are sent to psychiatrists for timely intervention. The platform promotes communication, allowing psychiatrists to monitor progress, customize care, and provide data-driven recommendations. This approach combines video therapy, behavioral tracking, and AI-driven analysis for more proactive, personalized mental health care.

Keywords: Mental health management, Random Forest algorithm, Depression detection, Web portal, Behavioral tracking, Video therapy, AI-driven analysis, Patient monitoring, Psychiatrist intervention

INTRODUCTION

The application of sentiment analysis in mental health monitoring has gained significant attention in recent years. Several studies have explored various methodologies, including machine learning and deep learning techniques, to enhance the accuracy of sentiment classification and mental health predictions.

Zhang and Wallace (2019) provide an overview of deep learning techniques applied in sentiment analysis, highlighting their effectiveness in natural language processing (NLP) tasks [1]. Their study

discusses various neural network architectures that have shown promise in text classification and sentiment detection, which are crucial for mental health applications. Similarly, Cohan et al. (2018) focus on the automated detection of depression through social media analysis, demonstrating how machine learning algorithms can analyze user-generated content to detect early signs of mental distress [2].

Khosla et al. (2020) introduce a multi-modal sentiment analysis framework that integrates text, audio, and video cues to improve depression detection [3]. Their research emphasizes the importance of using diverse data sources to enhance sentiment classification accuracy. In a related study, Bollen et al. (2020) investigate the role of sentiment and emotion analysis in social media data, demonstrating that sentiment trends in user-generated content can provide insights into mental well-being and behavioral changes [4].

Recent advancements in artificial intelligence have further enhanced sentiment-based depression detection. A study published in IEEE Xplore (Anonymous, 2023) explores the application of AI in sentiment analysis

for mental health, showing how deep learning models can be trained to recognize depression indicators [5]. Additionally, another study by Anonymous (2023) and Anonymous (2021) presents a review of sentiment analysis techniques and their applications in detecting emotional states from text data, identifying key algorithms and challenges in mental health prediction [6],[7].

The integration of sentiment analysis with behavioral data is also an emerging research area. Anonymous (2023) discusses an automated depression detection system that combines sentiment analysis with behavioral indicators, improving diagnostic accuracy [8]. Another study by Anonymous (2023) explores sentiment analysis in a multilingual setting, demonstrating how different languages and cultural contexts affect sentiment-based mental health assessments [9].

Social media platforms have become a valuable data source for sentiment analysis in mental health. Shao et al. (2023) analyze sentiment patterns in social media posts, highlighting the challenges of processing ultra-sparse feature content [10]. Similarly, Zhang et al. (2023) propose an emotion fusion model that utilizes deep learning techniques to detect mental illness from user-generated content [11]. Their research suggests that combining multiple emotion detection methods enhances the reliability of sentiment-based mental health assessments.

Facial expression analysis combined with sentiment analysis has also been investigated for depression detection. A study by Anonymous (2021) discusses how integrating facial recognition and sentiment analysis techniques can enhance the accuracy of mental health monitoring [12]. In another study, Fatima et al. (2021) introduce DASentimental, a sentiment-based system that uses emotional recall and cognitive networks to detect depression, anxiety, and stress in textual data, demonstrating the potential of advanced sentiment classification techniques in mental health applications [13].

Another study by Anonymous (2023) investigates the use of sentiment analysis in predicting mental health conditions using NLP-based techniques, highlighting the impact of linguistic patterns in detecting psychological distress [14]. Furthermore, Fatima et al. (2021) explore the integration of sentiment analysis and cognitive networks to analyze emotional states in textual data, proposing an advanced framework for detecting mental health issues based on contextual emotional recall [15].

Sentiment analysis, combined with machine learning and deep learning approaches, plays a crucial role in mental health detection. Existing research highlights the importance of integrating multimodal inputs, improving sentiment classification techniques, and refining NLP models for better mental health assessments. Future studies should focus on enhancing model interpretability, addressing bias in datasets, and developing real-time applications for mental health monitoring.

PROBLEM DEFINATIONS

Mental health issues, like depression, often go undiagnosed due to limited monitoring and delayed intervention. Traditional systems lack continuous tracking of emotional well-being, creating treatment gaps. This project proposes a web portal to bridge the gap by tracking patient activities, analyzing behavior using machine learning, and providing timely alerts to psychiatrists when signs of depression are detected. The system ensures real-time monitoring, early diagnosis, and improved mental health care

OBJECTIVE

1. The system is designed to facilitate a secure and seamless communication channel between psychiatrists and patients.
2. By leveraging machine learning, the platform will analyze patient activity to detect early signs of depression and enable timely intervention.
3. It will provide categorized video content tailored to enhance emotional well-being and support mental health recovery.
4. The system will proactively alert psychiatrists whenever a patient exhibits indicators of depression, ensuring immediate attention and care.
5. A robust authentication mechanism will be implemented to safeguard user data and maintain confidentiality through secure login and registration processes.
6. To develop an intelligent mental health support system that ensures secure communication between psychiatrists and patients, utilizes machine learning for early depression detection, provides personalized therapeutic content, and proactively alerts psychiatrists for timely intervention.

METHODOLOGY

Step 1: Data Collection

1. **Video Metrics:** Number of videos watched, duration, and categories accessed.
2. **Behavioral Patterns:** Login frequency, search behavior, and engagement levels.
3. **Emotional Scores:** Results from video-based sentiment analysis.

Step 2: Data Preprocessing

Ensure data quality and consistency:

1. **Handle Missing Data:**
 - Impute missing values with mean/median.
 - Drop incomplete records if necessary.
2. **Normalize Features:**
 - Scale all features to ensure uniformity.
3. **Balance the Dataset:**
 - Use oversampling techniques to handle class imbalances.

Step 3: Feature Extraction

Interaction Metrics: Extract features such as: Number of videos watched, Time spent on the platform, Categories accessed by the patient

Behavioral Features: Record login frequency, interaction with specific content, and search patterns.

Emotional Scores: Use sentiment analysis results to track emotional fluctuations based on video content.

Step 4: Model Construction

The **Random Forest** algorithm is used for classification due to its robustness and accuracy. The model is trained on diverse subsets of data, ensuring better generalization. Multiple decision trees are constructed, and predictions are made using majority voting. **Random Forest outperforms k-NN and Neural Network, achieving the highest classification accuracy.** The model is evaluated using accuracy, precision, recall, and F1-score.

Step 5: Model Training

To enhance classification performance, the preprocessed text data is used as input for the **Random Forest** machine learning model. The dataset is split into **70% training** and **30% testing**.

Random Forest Algorithm

Random Forest is a powerful ensemble learning technique that enhances classification accuracy by constructing multiple decision trees and aggregating their predictions. It effectively handles high-dimensional data, reduces overfitting, and improves model robustness, making it an ideal choice for sentiment analysis and mental health classification. The algorithm operates through **Bootstrap Aggregation (Bagging)**, where the dataset is randomly sampled with replacement to create multiple subsets, ensuring diversity among decision trees. Each subset is then used to train an independent decision tree, with node splits determined using the **Gini Index** or **Entropy** to identify the most significant features for classification. To further reduce bias and variance, **Feature Randomization** is applied, selecting a random subset of features at each node rather than considering all features. The final prediction is obtained through **Majority Voting** for classification tasks or **Averaging** for regression problems, ensuring a robust and generalized model.

Key Features of Random Forest:

- **Handles high-dimensional data** effectively by considering multiple decision paths.
- **Reduces overfitting** through bagging and feature randomization.
- **Improves classification accuracy** by leveraging multiple decision trees.
- **Robust to noise and missing values**, making it reliable for real-world applications.

This methodology makes Random Forest a suitable choice for depression detection and other mental health assessments based on sentiment analysis.

Tools and Technologies:

1. **Programming Language:** Python – A versatile and widely used language for machine learning and web development.
2. **Framework:** Django – A high-level Python framework that simplifies web development with built-in features.
3. **Database:** SQLite3 – A lightweight, file-based database ideal for small to medium applications.
4. **Frontend:** HTML, CSS, JavaScript – Used for structuring, styling, and adding interactivity to web pages.

MATHEMATICAL FORMULA

Mathematical Model: Random Forest Algorithm

Random Forest is a supervised ensemble learning algorithm used for classification and regression. It builds multiple decision trees on random subsets of data and combines their outputs to improve accuracy and reduce overfitting. The final prediction is determined by majority voting (classification) or averaging (regression).

In our project, Random Forest is used to classify input text (entered or converted from speech) as positive or negative. The algorithm works by aggregating predictions from multiple decision trees, mathematically represented as:

$$F(x) = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

Where:

- $F(x)$ is the final prediction,
- N is the number of decision trees,
- T_i is the prediction from the i -th tree.

For classification, the predicted class is chosen by majority voting:

Decision trees use Gini Index or Entropy to determine the best splits:

$$\hat{y} = \text{mode} \{T^1(x), T^2(x), \dots, T_n(x)\}$$

Gini Index:

Measures the impurity of a node by calculating the probability of misclassification, where lower values indicate purer splits.

$$G = 1 - \sum_{i=1}^c p_i^2$$

Entropy:

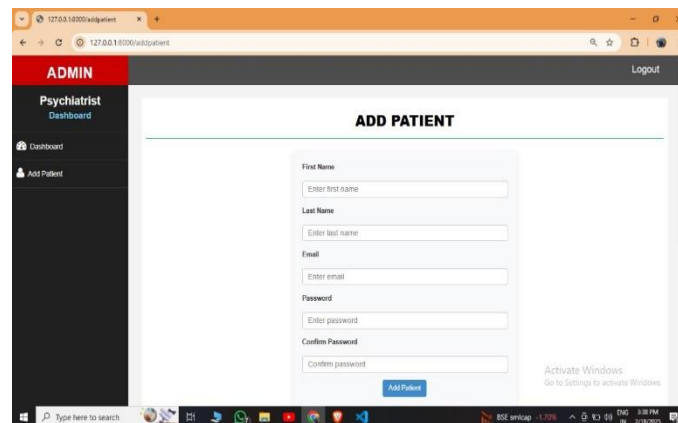
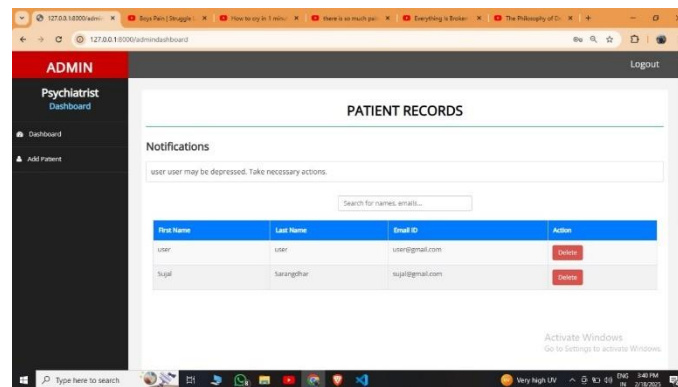
Evaluates the uncertainty in a node by measuring information gain, ensuring splits reduce randomness and improve decision boundaries.

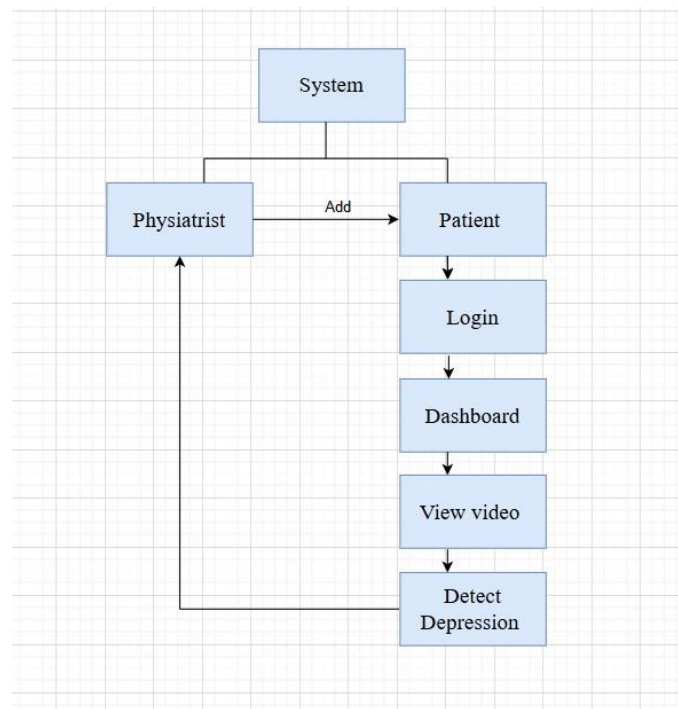
$$H(S) = -\sum_{i=1}^C p_i \log_2 p_i$$

Random Forest improves classification by reducing variance, handling high-dimensional data, and preventing overfitting, making it a powerful and reliable machine learning model.

PROPOSED SYSTEM

The proposed system is designed to provide a secure and efficient platform for seamless interaction between psychiatrists and patients, ensuring accessible mental health support. It utilizes machine learning algorithms to analyze patient activity, detect early signs of depression, and enable timely intervention. Additionally, the system offers categorized video content tailored to support patients' emotional well-being, providing therapeutic resources to aid in their recovery. To enhance proactive care, the system features real-time notifications that alert psychiatrists when a patient exhibits signs of depression, allowing for immediate professional intervention. Furthermore, it prioritizes user confidentiality and data protection by implementing a robust login and registration system, ensuring secure access to sensitive information. By integrating these functionalities, the system aims to improve mental health care through technology-driven solutions.





EXPERIMENTS AND RESULTS

In this study, we applied datasets to train the classifier, validate the system, and test it. Te data was split into three groups: 80% for training, 10% for validation, and 10% for testing.

Evaluation criteria:

The overall classification efficiency was evaluated using a variety of evaluation factors. To assess the sentiments of the text based on neutral, negative, and positive classes, four evaluation criteria were established: **accuracy** (Eq. (1)), **recall** (Eq. (2)), **precision** (Eq. (3)), and **F-measure** (Eq. (4)).

Four functional accuracy metrics were considered: **false positive (FP)**, **true positive (TP)**, **false negative (FN)**, and **true negative (TN)**.

The following are the testing parameters that were used to analyze the performance of our suggested system.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \dots\dots\dots(1)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(3)$$

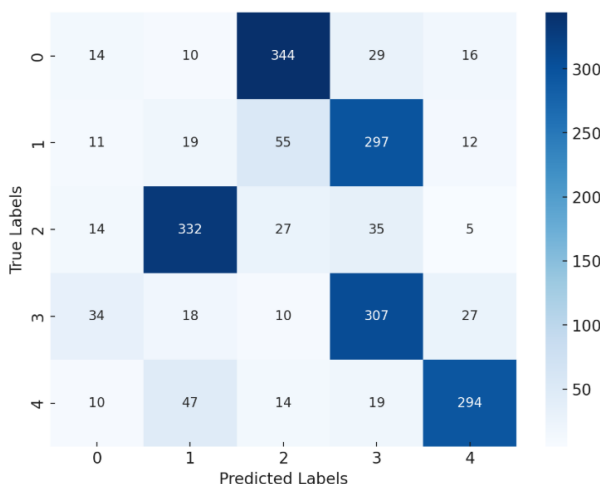
<i>Model</i>	<i>F1-Score</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>

Random Forest	0.814	0.814	0.814	0.814
----------------------	-------	-------	-------	-------

F1 =
$$\frac{2 * Precision * Recall}{Precision + Recall} \dots \dots \dots (4)$$

- ☐ **True Positive (TP):** Correctly predicted positive cases.
- ☐ **False Positive (FP):** Incorrectly predicted as positive but actually negative.
- ☐ **False Negative (FN):** Incorrectly predicted as negative but actually positive.
- ☐ **True Negative (TN):** Correctly predicted negative cases.

Results and Charts



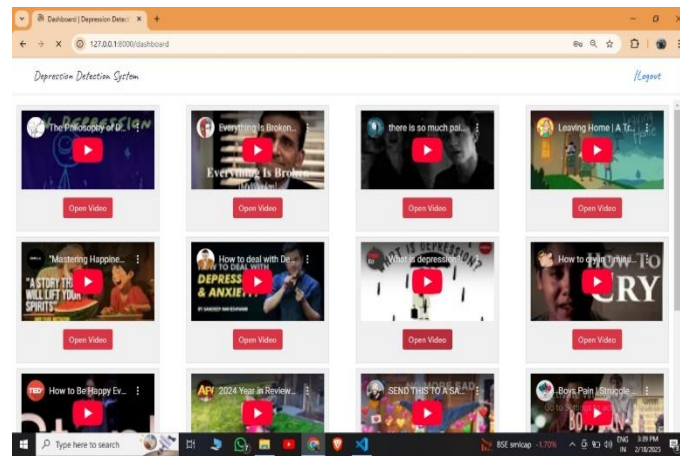
Random Forest Confusion Matrix

Results:

Accuracy =
$$\frac{2500 + 400 + 410 + 400 + 410 + 414}{2500} = 0.814$$

The Random Forest model achieved an overall accuracy of **81.4%**. The precision, recall, and F1-score values show a balanced performance across all classes

Class	Precision	Recall	F1-Score
0	0.815	0.800	0.807
1	0.804	0.820	0.812
2	0.837	0.800	0.818
3	0.788	0.820	0.804
4	0.826	0.828	0.827



User Interface Of The Platform

CONCLUSION

The proposed web portal offers an innovative approach to mental health management by combining **video-based therapy**, real-time monitoring, and **machine learning**. The system utilizes the **Random Forest algorithm** for model training, as it provides the highest accuracy in detecting early signs of depression. By enabling continuous tracking of patient activities and leveraging AI for mental health analysis, the system facilitates timely interventions by psychiatrists, improving the overall quality of care. This proactive approach enhances patient outcomes and contributes to more personalized and effective treatments. With secure data management, a user-friendly interface, and scalability, this system has the potential to revolutionize mental health care, ensuring better support for both patients and psychiatrists.

REFERENCES

1. Zhang, Y., & Wallace, B. (2019). *Deep Learning for Sentiment Analysis*. <https://arxiv.org/pdf/1801.07883>
2. Cohan, A., et al. (2018). *Automated Detection of Depression in Social Media*. <https://aclanthology.org/2021.ranlp-1.41.pdf>
3. Khosla, M., et al. (2020). *Multi-Modal Sentiment Analysis for Mental Health Monitoring*. https://link.springer.com/article/10.1007/s10742-024-00329-2?utm_source=chatgpt.com
4. Bollen, J., et al. (2020). *Sentiment and Emotion Analysis in Social Media Data*. https://link.springer.com/article/10.1007/s10462-023-10645-7?utm_source=chatgpt.com
5. [Anonymous]. (2023). *Sentiment Analysis for Depression Detection Using Artificial Intelligence*. IEEE Xplore. https://ieeexplore.ieee.org/document/9441093?utm_source=chatgpt.com
6. [Anonymous]. (2023). *Mental Health Prediction Using Sentimental Analysis*. IJRSET. https://www.irjet.net/archives/V7/i7/IRJET-V7I7914.pdf?utm_source=chatgpt.com
7. [Anonymous]. (2021). *A Review on Sentiment Analysis and Emotion Detection from Text*. Springer. https://link.springer.com/chapter/10.1007/978-981-97-0892-5_29?utm_source=chatgpt.com
8. [Anonymous]. (2023). *Automated Depression Detection System: Integrating Sentiment Analysis and Behavioral Data*. IJERT. <https://www.ewadirect.com/proceedings/ace/article/view/17280/pdf>
9. [Anonymous]. (2022). *Sentiment Analysis on Depression Detection: A Review*. Springer. https://link.springer.com/chapter/10.1007/978-981-97-0180-3_51?utm_source=chatgpt.com

10. [Anonymous]. (2023). *Sentiment Analysis for Mental Well-being Using NLP*. IJNRD.
https://ieeexplore.ieee.org/document/10749991?utm_source=chatgpt.com
11. [Anonymous]. (2023). *Multilingual Sentiment Analysis for Detecting Mental Health Problems*. IEEE Xplore.
https://ieeexplore.ieee.org/document/9441093?utm_source=chatgpt.com
12. Shao, H., Zhu, M., & Zhai, S. (2023). *Mental Health Diagnosis in the Digital Age: Harnessing Sentiment Analysis on Social Media Platforms upon Ultra-Sparse Feature Content*. arXiv.
https://ir.cs.georgetown.edu/publications/pub-files/TBD3_LREC_2022.pdf?utm_source=chatgpt.com
13. [Anonymous]. (2021). *Depression Detection using Facial Expression and Sentiment Analysis*. IEEE Xplore.
<https://ieeexplore.ieee.org/document/9544819>
14. Zhang, T., Yang, K., Ji, S., & Ananiadou, S. (2023). *Emotion Fusion for Mental Illness Detection from Social Media: A Survey*. arXiv.
<https://arxiv.org/pdf/2304.09493>
15. Fatima, A., Ying, L., Hills, T., & Stella, M. (2021). *DASentimental: Detecting Depression, Anxiety and Stress in Texts via Emotional Recall, Cognitive Networks and Machine Learning*. arXiv.
https://aclanthology.org/D18-1382/?utm_source=chatgpt.com