IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

LEVERAGING AI TO PREDICT MENTAL ILLNESS FROM CURRENT LIFESTYLE AND SITUATIONAL DATA

[1]Fariha Baqi, [2]Mr. Aaftab Alam, [3]Dr. Mohd Haroon

[1]M. Tech Scholar, [2]Assistant Professor, [3]Professor
[1]Department of Computer Science & Engineering,
[1]Integral University, Lucknow, India

Abstract: In recent lots of people are experiencing depression, anxiety, or stress and these conditions really damage people's health.The problem is well recognized, but most individuals do not receive a proper diagnosis or treatment due to the social stigma associated with mental disorders, lack of mental health services, and our traditional method of diagnosing diseases, which is dependent upon self-reporting. Luckily, artificial intelligence is improving rapidly and because we have so much digital information, we can now create new algorithms based on data to identify mental health problems.This project is about using artificial intelligence to find mental disorders and it does this by looking at both organised and less organised information using machine learning. In order to make things clear, it uses traditional approaches such as Logistic Regression, SVM, and Random Forest; as well as sophisticated deep learning models such as LSTM and BERT. The older approaches are good because they are easy to understand and work brilliantly with straightforward information, while Deep Learning models are for finding hidden patterns and using those to predict what's going on. The system uses a chatbot to get information as it's needed. Accuracy-wise, the proposed hybrid approach shows excellent performance accuracy, having an accuracy prediction rate of 92.50% with the AUC-ROC score of 0.9759. This means it's more accurate than any of the models used on their own, but it is just as good at being precise, and at finding all the relevant information.

KEYWORDS - Artificial Intelligence, Mental Health Prediction, Machine Learning, NLP, Chatbot, Lifestyle Data, Digital Psychiatry

1. INTRODUCTION

Mental wellbeing is another important element that will influence the level of wellness of the person as it will determine the manner in which the individual will react to different circumstances [9], [16]. Nowadays, disorders like depression, anxiety, and stress have been observed to be rising rapidly [10], [15]. They are one of the main causes of disability across the globe and have a huge impact on the quality of people's lives [9], [10]. The reason why individuals fail to receive treatment for their psychological conditions is due to numerous factors including the stigma attached to mental illness, lack of professionals and institutions offering services for such conditions, and the current diagnostic process of mental disorders which relies on interviews and surveys [1], [31]. Conventional techniques that have been used in diagnosing mental problems cannot be scalable and are subject to the subjective information obtained from the patient [1], [2]. Therefore, technology must be applied to develop tools that will be useful in diagnosing mental problems[3], [4].

Herein, having access to a large volume of data along with the advent of technology associated with AI has made research focused on harnessing the capabilities of data science for the prediction of mental diseases [2], [21]. Scientists can use machine learning algorithms to analyze a large amount of data, such as social media data, lifestyle data, text input, etc [2], [7], [24]. It allows them to understand the intricate psychological and behavioral aspects of mental well-being [15], [20]. Also, it has been found that text data in particular is very helpful in understanding the psychological state of an individual based on his language [7], [22].

The logistic regression, support vector machine, and random forest algorithms belong to the most commonly employed machine learning algorithms to perform the classification and prediction of mental well-being [2], [5]. The estimation of the probability of an event happening is made using logistic regression with the help of the Sigmoid function [2]:

$$P\left(y=\frac{1}{x}\right)=\frac{1}{1+e^{-(z)}} \qquad (1.1)$$

where $z = w^T x + b$, representing a linear combination of input features.

$P(y = 1 \mid x)$ → Probability of class 1 , x → Input features , w → Weight vector , b → Bias , e → Euler's constant.

SVM is used to classify users into different mental health risk levels based on questionnaire and behavioral features. SVM identifies an optimal hyperplane that maximizes the margin between classes [2], [5]:

$$min\frac{1}{2}\|w\|^2 \; subject\ to\ y_i(w^T x_i + b) \geq 1 \qquad (1.2)$$

where $w^T x + b = 0$wTx+b=0 represents the optimal separating hyperplane.
$w \rightarrow$ Weight vector defining the hyperplane, $b \rightarrow$ Bias term, $x_i \rightarrow$ Input feature vector, $y_i \rightarrow$ Class label (+1 or −1), $\|w\| \rightarrow$ Magnitude of weight vector (controls margin width).
It is effective for high-dimensional data and improves classification performance [5], [11]. Instead of just one combination, Random Forest builds lots of decision trees and then puts all their predictions together to get a better one [2], [9].

$$\hat{y}\frac{1}{N}\sum_{i=1}^{N} h_i(x) \qquad (1.3)$$

where each $h_i(x)$ represents the prediction from the i^{th} decision tree in the Random Forest.
$\hat{y} \rightarrow$ Final predicted output, $N \rightarrow$ Total number of decision trees, $h_i(x) \rightarrow$ Prediction of the i^{th} tree, $x \rightarrow$ Input feature vector
They work well with organized data and it's fairly easy to understand why they make a certain prediction, but you usually have to work on the data yourself to decide which characteristics are important [9], [26]. They do not function as well with complicated data that has multiple variables or sequences [5], [20]. For instance, they cannot work properly with textual data where the order is important [22], [24]. One solution to this problem can come from the application of deep learning models [5], [15]. LSTM neural network models have proven to be effective for dealing with sequence data and finding patterns of information over time in text and behavioral data [5], [13]. In addition to this, the use of techniques such as BERT for instance has transformed the domain of natural language processing through the implementation of bi-directional attention modeling as follows [6], [28]:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1.4)$$

where the attention mechanism computes the relevance between input tokens by comparing query and key vectors and scaling the result.
$Q \rightarrow$ Query matrix, $K \rightarrow$ Key matrix, $V \rightarrow$ Value matrix, $K^T \rightarrow$ Transpose of key matrix, $d_k \rightarrow$ Dimension of key vectors (used for scaling), $softmax \rightarrow$ Normalizes scores into probabilities
Despite their effectiveness, these models are often applied independently, which limits their ability to leverage complementary strengths. The methodologies have proven to be very promising in terms of detecting various psychological diseases using textual analysis, specifically by taking into account social media conversations [2], [7], [23]. Yet, the emerging tendency of digitizing psychiatry has been seen by many people as creating applications or even AI chatbots that can help the patient monitor their own health [6], [12], [29]. These bots can communicate instantly, thereby making it easy for doctors to communicate with their patients instantly [7], [12]. The problem is that the current technology in place can do any one but not both tasks at once [16], [31].
Moreover, there are certain factors that can create difficulties in adopting such AI-driven systems to promote mental health, which include things like differences in the origin of data, biases within self-reported information, and understanding the model [3], [10]. It is, therefore, clear that there is need for an all-encompassing approach [14], [26]. In this regard, this current research paper presents an advanced approach to AI technology by making use of not only machine learning methods such as logistic regression, random forest, and SVM but also deep learning methods such as LSTM and BERT, along with a conversational chatbot [2], [5], [6]. By incorporating lifestyle, situational, and textual factors, the proposed system seeks to enhance early diagnosis and tracking of mental illnesses while also enhancing usability [3], [10], [21].

2. OBJECTIVE OF THE STUDY

The purpose of this research is to develop a complete, automated, and artificial intelligence AI-based solution consisting of machine learning-based predictive models and a conversational chatbot to achieve early detection of, and ongoing assessment of mental health disorders.

3. LITERATURE REVIEW

Table 1: Comprehensive Literature Review

S. No.	Research Paper Details	Methodology	Results and Discussion	Findings	AUC-ROC	Research Gap	Future Scope
1	Reece & Danforth (2017) – *Instagram Photos Reveal Predictive Markers of Depression*	ML-based Instagram image analysis using color, brightness, facial detection,	Machine learning models outperform traditional clinical assessment and enable early depression detection	Social media–based visual and behavioral data can serve as early indicators of depression	Not explicitly reported	Image-only approach; excludes lifestyle context and user interaction	Integrate visual data with lifestyle and conversational inputs
2	Fitzpatrick et al. (2017) – *Woebot: Conversational Agent for CBT*	NLP-based CBT chatbot evaluated through a randomized controlled trial on students	Reduced depression and anxiety with high user engagement	Chatbots provide scalable and accessible mental health support	Outcome-based evaluation	Lacks predictive capability; focused only on therapy	Integrate chatbot with AI-based risk prediction models

3	Fulmer et al. (2018) – Psychological AI (Tess) for Depression & Anxiety (JMIR Mental Health)	RCT with college students evaluating AI chatbot (Tess) using PHQ-9, GAD-7, and PANAS	Reduced depression and anxiety with high satisfaction and strong engagement	AI chatbot effectively provides scalable and cost-efficient mental health support	Clinical outcome-based	Short-term study with limited scope and no predictive or multimodal analysis	Incorporate long-term evaluation, predictive analytics, and multimodal personalization
4	Nichols et al. (2016) – *Depression Prediction Using Primary Care Records*	Logistic regression on large-scale EHR data including psychological and socioeconomic variables	Achieved AUC ≈ 0. 72; suitable for clinical screening	AI supports early mental health screening in healthcare systems	~0. 72	Limited to clinical data; lacks real-time and lifestyle context	Integrate EHR with lifestyle and behavioral data for proactive monitoring
5	Vaidyam et al. (2019) – Review of mental health chatbots (Canadian Journal of Psychiatry)	Systematic review of chatbot applications across multiple mental health conditions	Effective for CBT, psychoeducation, and symptom monitoring with high user satisfaction	Chatbots enhance accessibility and user comfort in mental healthcare	Conceptual study	Lacks standard evaluation, long-term validation, and integration with predictive models	Develop standardized metrics and integrate chatbots with AI-based lifestyle prediction systems
6	Rahman et al. (2020) – *Systematic Review of ML for Mental Health Detection*	Review of ML/NLP models using social media and online text data	SVM, RF, and NB achieve high accuracy in detecting depression and suicide risk	NLP-based ML is effective for large-scale mental health analysis	0. 88–0. 96	Limited to static text analysis; lacks personalization and context awareness	Develop multimodal models incorporating behavioral and lifestyle data
7	Abd-Alrazaq et al. (2020) – *Meta-analysis of Mental Health Chatbots*	Systematic review and meta-analysis of chatbot interventions	Shows moderate improvement in depression and stress	Chatbots are promising tools for mental health support	Outcome-based evaluation	Lacks long-term validation and predictive capability	Integrate chatbots with continuous risk monitoring systems
8	Abd-Alrazaq et al. (2022) – *The performance of artificial intelligence-driven technologies in diagnosing mental disorders: An umbrella review* (npj Digital Medicine)	Umbrella review of multiple AI studies using clinical and neuroimaging data	AI models show wide accuracy variation with high performance in neuroimaging-based studies	AI enables early and objective diagnosis, influenced by data type and validation	0. 75–0. 99	Focused on clinical data; lacks lifestyle context and generalizability	Incorporate lifestyle data and develop explainable, real-time AI systems

9	Bond et al. (2023) – *Digital Transformation of Mental Health Services* (npj Mental Health Research)	Framework-based study analyzing digital mental health technologies across disciplines	Digital tools enhance care through monitoring, real-world data, and integrated services	Digital technologies improve accessibility, personalization, and continuity of care	Conceptual Study	Lacks predictive models and integrated AI-based diagnostic systems	Develop AI-driven predictive systems with real-time interaction and ethical governance
10	Iyortsuun et al. (2023) – *ML & DL for Mental Health Diagnosis*	PRISMA-based review of ML and DL models including CNN, LSTM, SVM, and RF	Deep learning enhances diagnostic accuracy across multiple data types	DL is effective for analyzing complex mental health data	0. 88–0. 99	Interpretability and ethical challenges remain unresolved	Emphasize development of explainable and ethical AI systems
11	Zakariah & Alotaibi (2023) – *Actigraphy-Based Depression Detection*	Neural network analysis of wearable sensor data (motor activity)	Achieves up to 99% accuracy in depression classification	Wearable data enables objective mental health monitoring	~0. 99	Lacks conversational and psychological context	Integrate wearable data with chatbot-based assessment
12	**Squires et al. (2023)** – Deep Learning & ML in Psychiatry (Brain Informatics)	Comprehensive survey of ML, DL, NLP, and multimodal approaches for mental health analysis	AI supports detection, diagnosis, and treatment prediction using diverse techniques	Multimodal AI enables personalized psychiatry and improves predictive performance	Conceptual analysis	Limited validation, small datasets, and lack of standardized evaluation	Develop large-scale, validated, and explainable AI systems with standardized frameworks
13	Ku & Min (2024) – ML Stability in Predicting Depression & Anxiety (Healthcare)	Compared five ML models (CNN, Random Forest, XGBoost, Logistic Regression, Naïve Bayes) using survey and EHR data with noise-based robustness testing	Performance declines with noisy data; CNN shows highest robustness	CNN demonstrates strong resilience in handling subjective data	High (not specified)	Focuses on robustness; lacks real-time, personalized, and interactive systems	Develop real-time, explainable AI systems integrating conversational and lifestyle data
14	Zafar et al. (2024) – *AI for Depression and Anxiety Detection*	Review of NLP, emotion recognition, and deep learning models	AI enables effective early risk assessment and diagnosis	Multimodal AI enhances mental health prediction	Not explicitly reported	Lacks deployment-ready integrated systems	Develop end-to-end AI-driven mental healthcare platforms
15	Olawade et al. (2024) – *Enhancing Mental Health with AI*	Narrative review of AI applications in teletherapy and diagnosis	AI supports personalized treatment and virtual therapy	AI improves accessibility and helps reduce stigma	Conceptual study	Ethical, regulatory, and trust challenges remain	Develop ethical and regulated AI frameworks

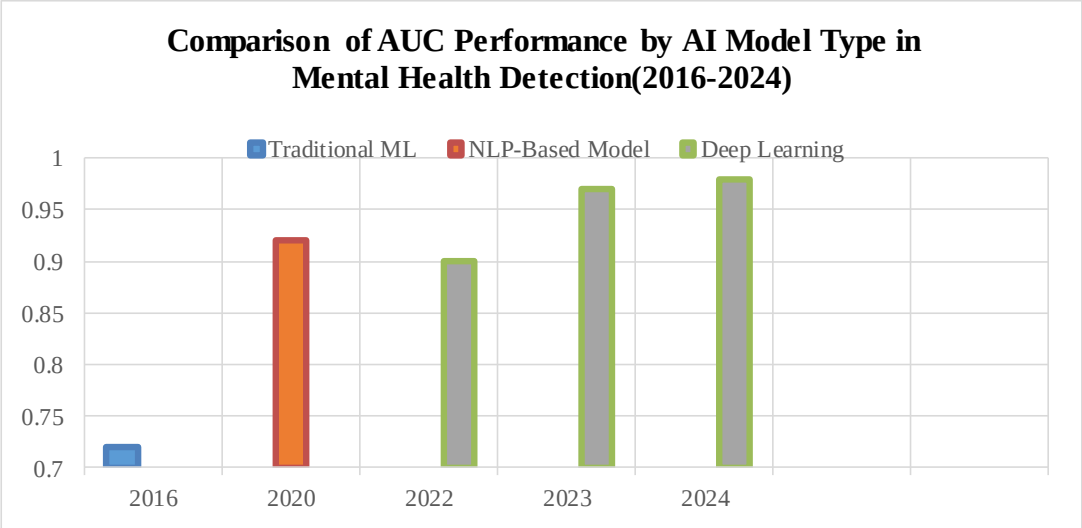


Fig 1. Comparison of AUC-ROC performance across different AI model types for mental health detection.

3.1 RESEARCH GAP

Since 2017, there has been an increase in the use of AI and ML for providing support for mental health and finding/diagnosing illnesses using this technology. To resolve the previously mentioned problems, this proposed work will develop a multimodal and ensemble-based methodology by combining clinical data with textual data to enhance the predictive accuracy of mental health problem model sand to improve the interpretability of such models. In addition, from 2023 to 2024, research focused on using deep learning techniques with wearable devices to increase accuracy when predicting outcomes in individuals with mental health problems. Most of the existing studies currently use one mode of data and do not combine clinical data with textual data. When relying only on one mode of data, these studies have difficulty capturing the full context of mental health conditions. Moreover, the traditional machine learning models lack the capability to capture intricate patterns, while the modern modeling approaches, such as BERT and LSTM models, are generally applied independently rather than together. Besides, the use of ensemble modeling and multi-modal feature extraction is not common, hence compromising the efficacy of the models developed. In light of solving the aforementioned issues, the proposed research shall develop an ensemble model through integration of textual and clinical data for improving the effectiveness of models predicting mental health problems.

4. RESEARCH METHODOLOGY

4. 1 PROPOSED RESEARCH FRAMEWORK

This study introduces a framework for the early prediction of mental health conditions using a combination of artificial intelligence techniques with both structured clinical and social media data and unstructured textual data. Machine Learning techniques are used in combination with Deep Learning techniques to enhance the accuracy of the prediction results. The proposed framework uses the multimodal approach in which the text data is processed by using the BERT-based LSTM model, and the structured clinical data is normalized separately. The output from the LSTM model is then used in the machine learning model.

4. 2 DATASET DESCRIPTION

The dataset used consists of approximately 14,796 different samples (1. 64GB) that has been gathered for this research project. The primary source of the training data used by the model is textual data. The textual data is obtained from different social media platforms such as Reddit. The dataset is publicaly available at :

Dataset: *https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned*

The data obtained from clinical questionnaires such as PHQ-9 and GAD-7 are not used for training the model. Instead, it is provided as additional features to the model to improve the accuracy of the predictions obtained from the model.

The dataset is represented as:

$$D = \{(x_i, y_i)\} \text{ for } i = 1 \text{ to } n \qquad (4.1)$$

where each pair (x_i, y_i) represents a single data instance in the dataset.

$D \rightarrow$ Dataset, $x_i \rightarrow$ Input feature vector of the i^{th} sample, $y_i \rightarrow$ Corresponding output label (class), $n \rightarrow$ Total number of samples where x_i represents the feature vector and y_i represents the corresponding mental health class. The model can then learn to associate each of these features with the various mental health conditions represented in each of the samples during model training.

4. 3 DATA PREPROCESSING

Data preprocessing is performed to ensure that the data is of good quality. This is done by handling missing values using imputing techniques. Normalization is done using standardization for numeric values.

$$Z = \frac{X - \mu}{\sigma} \qquad (4.2)$$

where standardization transforms the data to have zero mean and unit variance.

$Z \rightarrow$ Standardized value, $X \rightarrow$ Original feature value, $\mu \rightarrow$ Mean of the feature, $\sigma \rightarrow$ Standard deviation of the feature

Data preprocessing techniques used for unstructured data includes tokenization and sentiment analysis followed by tokenization using a pre-trained BERT tokenizer to generate contextual embeddings. The input data is processed in its raw format while being executed. Splitting of data is done in the ratio of 80:20 for training and testing.

4. 4 FEATURE ENGINEERING AND SELECTION

Feature engineering will first create context embedding using BERT, then generate sequences using LSTM. The resulting clinical scores will then have to go through normalization before being included with the outputs from LSTM to create a single vector of features. Along with feature engineering, there are many different ways to do feature selection as well. These techniques assist in the removal of redundant features. Features are selected based on their prediction accuracy, efficiency, and generalization.

4. 5 HYBRID PREDICTION STRATEGY (PROPOSED CONTRIBUTION)

The proposed system integrates outputs from ML and DL models using a Hybrid approach:

$$Final\ Prediction = \alpha \cdot ML_{output} + \beta \cdot DL_{output} \qquad (4.3)$$

where the final prediction is obtained by combining outputs from machine learning and deep learning models using weighted aggregation.

$\hat{y}$ → Final predicted output, y_{ML} → Prediction from machine learning models, y_{DL} → Prediction from deep learning models, α → Weight assigned to ML output, β → Weight assigned to DL output

Hybrid Ensemble Framework is one of the ways in which the predictions of various machine learning and deep learning algorithms can be combined into a single prediction. In the case of the LSTM model, it is used in conjunction with the set of predictions of the classification model. The predictions are then used in the making of the final prediction. In order for it to be more reliable in its usage, we are proposing the inclusion of a rule-based override facility. More weightage is assigned to the clinical parameters based on the PHQ-9 and GAD-7 scores crossing certain cutoff points. This improves accuracy, reduces bias, and enhances generalization.

4. 6 PREDICTION PROCESS

A typical prediction workflow includes importing the data, cleaning and getting it ready, identifying the most useful features, training the model, and then running the final classification. In these models, the system uses machine learning and deep learning to make predictions. After that, we use a mix of methods to combine the predictions from the different models. In the end, you get a risk score, and it's used to sort users into low-, moderate-, or high-risk groups.

4. 7 IMPLEMENTATION OVERVIEW

The coding of the system is carried out using Python, and it incorporates tools like Pandas, NumPy, and Scikit-learn, making it easier to process the data and make predictions in real time when running the system. We are also able to adjust settings for the smooth running of the system efficiently.

4. 8 PERFORMANCE EVALUATION

Model performance is evaluated using standard classification metrics:

4. 8. 1 ACCURACY

It measures the overall correctness of the model by calculating the proportion of correctly predicted instances (both positive and negative) out of all predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4.4)$$

where Accuracy measures the overall correctness of the model by calculating the proportion of correctly predicted instances.

TP → True Positives, TN → True Negatives, FP → False Positives, FN → False Negatives

4. 8. 2 PRECISION

This indicates the proportion of correctly predicted positive cases out of all predicted positive cases, reflecting how accurate the model is when it predicts a positive class.

$$Precision = \frac{TP}{TP+FP} \qquad (4.5)$$

where Precision indicates the proportion of correctly predicted positive cases among all predicted positive cases.

4. 8. 3 RECALL(SENSITIVITY)

It measures the proportion of actual positive cases that are correctly identified by the model, showing its ability to detect true positives.

$$Recall = \frac{TP}{TP+FN} \qquad (4.6)$$

where Recall measures the proportion of actual positive cases correctly identified by the model.

4. 8. 4 F1-SCORE

It is the balanced mean of precision and recall, providing a balanced measure when there is an uneven class distribution.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (4.7)$$

where F1-Score represents the harmonic mean of precision and recall, providing a balanced evaluation metric, especially for imbalanced datasets.

Precision → Correctness of positive predictions, Recall → Ability to detect actual positives

4. 8. 5 AUC-ROC

It is used to evaluate model performance across different thresholds. These metrics provide a comprehensive assessment of model effectiveness.

4. 9 MODEL INTERPRETABILITY

Interpretability of the model is also a part of this, and the aim here is to make the predictions of the model more interpretable and, in turn, make people feel more comfortable trusting the predictions of the model. Explainability is implemented using a model-agnostic tool which is SHAP used to interpret the model and make sense of the predictions made by the model.

5. SYSTEM DESIGN AND IMPLEMENTATION

5. 1 PROPOSED SYSTEM ARCHITECTURE

This new approach will leverage advanced Artificial Intelligence techniques to provide a framework for the proposed system to be developed on a multimodal basis, developing, processing, and delivering textual data through a BERT-Based LSTM architecture. Clinical information such as PHQ-9 and GAD-7 are normalized and used as structured features. The probability of output produced by the proposed model will be combined with clinical information and is used as input for machine learning classifiers such as Logistic Regression, Random Forest, and Support Vector Machine (SVM). A Hybrid mechanism will be developed to fuse the output generated by all the models to generate the final output, which is used to classify the user's information as low, moderate, or high risk and will also be used by the proposed chatbot system.

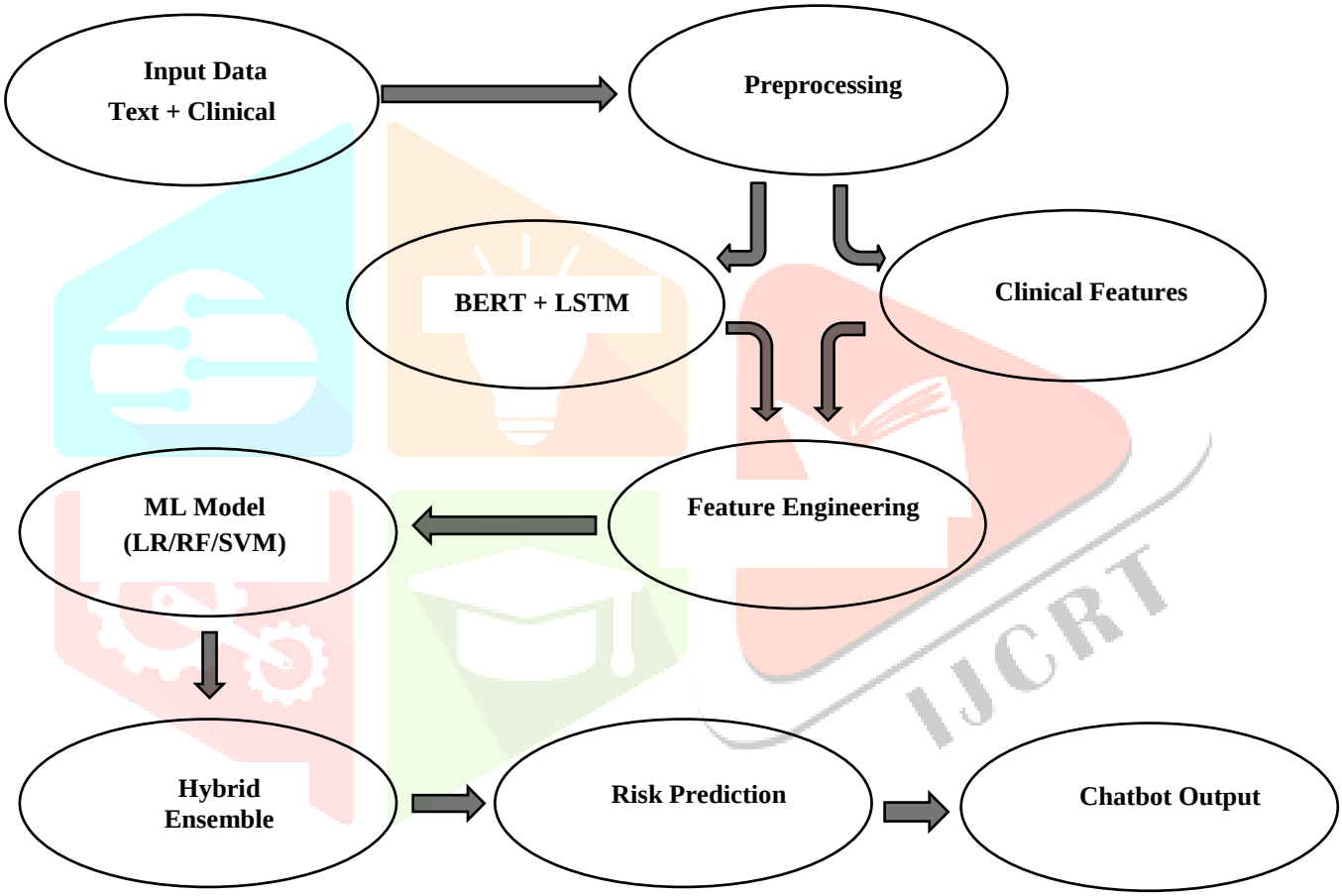


Fig 2: Simplified architecture of the proposed multimodal mental health prediction system

5. 2 IMPLEMENTATION DETAILS

With the help of algorithms that include machine learning and natural language processing, all done with the help of the Python language. For machine learning, we have some algorithms such as Logistic Regression, Random Forest, and Support Vector Machines(SVM). All these are available with the Scikit-learn library. For deep learning, we have the PyTorch library. For the natural language processing part, we have the BERT model to help us understand the context with the help of the data we are training. Before we do all this, during the initial stages, we make sure that all the information is available and all the numbers are scaled properly. After this, the whole set of data is split into two sets: one to train the machine learning model and another to test the machine learning model. We create the machine learning model with the help of something called supervised learning. We make some changes to make the machine learning model more accurate, called hyperparameters. After all this, the machine learning model will provide a score for all the users to classify their mental health risk into different categories.

5. 3 EXPLAINABILITY INTEGRATION

The system described will also have an explainability component to provide transparency for users by implementing SHAP (SHapley Additive exPlanations) explainability methods which will help identify which aspects are influencing the predicted outcome such as responses to questions, emotional responses and behaviors. These methods will provide a global and local understanding of the result of the prediction. The system's ability to be easily understood by the user will further provide reliability and be user friendly.

5. 4 SYSTEM OUTPUT

The system divides users into three categories (low, moderate, high risk) as well as provides continuous analysis of user input for real-time monitoring. After predicting a user's risk level, the system will provide the user with individualized recommendations for stress relief, lifestyle improvements and how to find a professional who can assist them. The chatbot creates ongoing interactions by tracking mood, gathering data about the user, and having conversational support with the user, which increases user interaction with the system and makes the system more effective.

6. RESULTS AND ANALYSIS

This section gives a general overview of the proposed system and how it predicts mental health outcomes. In this section, we will go over how all the machine learning models performed, then look at the deep learning models, and finally discuss the Hybrid ensemble model, using common evaluation measures like accuracy, precision, recall, F1 score, and the area under the ROC curve, and then compare it with other approaches to evaluate the effectiveness of the proposed method. Also, the paper discusses the explainability results and compares them with earlier studies, so it's easier to see what works well in the proposed approach and what feels new about it.

6. 1 PERFORMANCE OF INDIVIDUAL MODELS

The system was evaluated using four baseline models: Logistic Regression, Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM). The experimental results are summarized in Table 2. Their performance was compared against the proposed Mental Health Prediction, Hybrid Ensemble model.

Table 2: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
LSTM (Deep Learning)	92. 80%	0. 93	0. 93	0. 93	0. 9785
Logistic Regression	89. 32%	0. 89	0. 89	0. 89	0. 9543
Random Forest	83. 21%	0. 83	0. 84	0. 83	0. 9216
SVM	90. 24%	0. 90	0. 90	0. 90	0. 9606
Hybrid Ensemble	**92. 50%**	**0. 92**	**0. 93**	**0. 92**	**0. 9759**

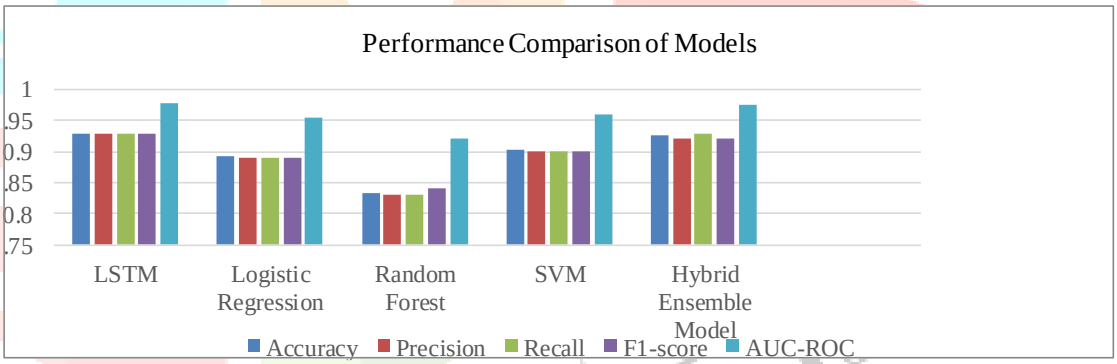
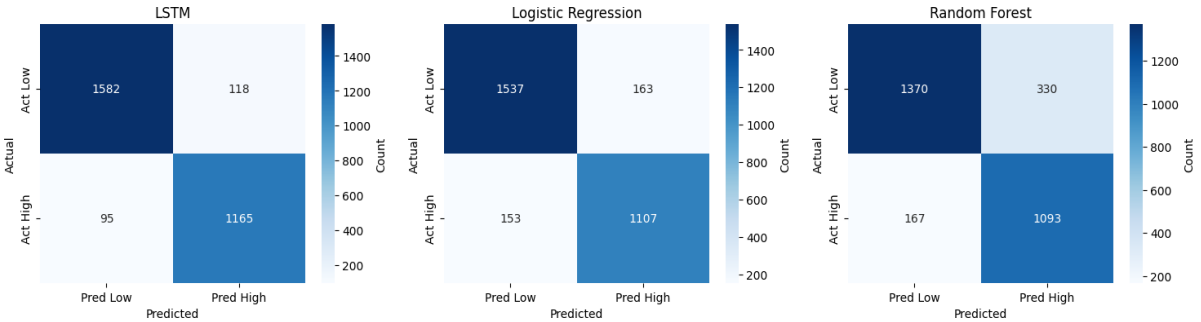


Fig 3: Performance Comparison of Models for Mental Health Prediction

According to comparison table, we can conclude that the LSTM performed better than the more conventional classifiers, such as SVM, Random Forest, Logistic Regression, etc. This implies that the model is better at recognizing more complex patterns in the sequence of text. The Hybrid ensemble performed better in terms of consistency with regard to all the metrics. This is because the predictions were more balanced compared to what any of the models would have performed.

6. 2 CONFUSION MATRIX

The outcomes being derived from the analysis using the confusion matrix revealed that the Hybrid Ensemble Model had the highest true positive and true negative values, 1177 and 1561 respectively. Moreover, this model had the fewest errors because it had the least number of false negatives, which was 83. This model is quite important in detecting people who are affected by mental disorders. We have another excellent model which is LSTM since it made the fewest mistakes. Nonetheless, all the other machine learning models did not perform well with respect to mistakes compared to the Hybrid Ensemble Model.



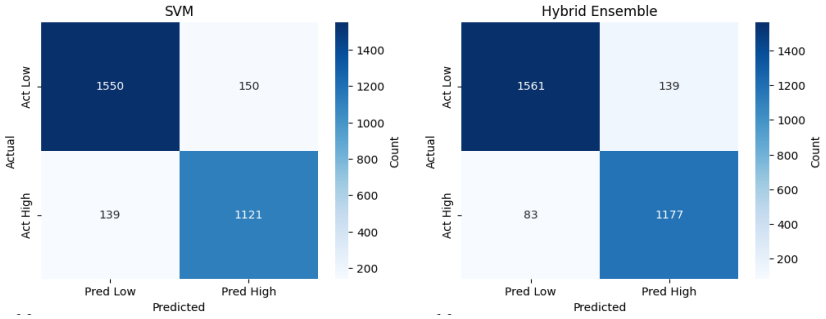


Fig 4: Confusion matrices of LSTM, Logistic Regression, Random Forest, SVM, and Hybrid Ensemble models, illustrating classification performance across low-risk and high-risk classes.

6. 3 PERFORMANCE OF THE HYBRID ENSEMBLE MODEL

The Hybrid Ensemble Model has very accurate measures at 92. 50% accuracy, AUC-ROC of 0. 9759, precision of 0. 92, recall of 0. 93, and F1-score of 0. 92. From the Hybrid Ensemble Model, it shows a balanced system where there is no imbalance between the two classes, while the positive class shows high measures of precision and recall, meaning this model is highly effective in detecting high-risk individuals without ignoring positives.

6. 4 EXPLAINABILITY RESULTS

If SHAP (SHapley Additive exPlanations) analysis is applied, we can obtain more information on how the features affect predictions for particular inputs. Positive SHAP values indicate features that increase the probability of high-risk predictions, while negative values indicate features that decrease it. Thus, SHAP could be useful in figuring out which features contribute positively and negatively to high-risk predictions of the Hybrid Ensemble model (textual features, structured data, and LSTM). Additionally, SHAP analysis is performed to understand the total contribution of the features toward high-risk predictions in the context of the Hybrid Ensemble model. Because of the use of SHAP in the analysis of the dataset, we can identify the following as key predictors: anxiety, depression, stress, sadness, and the first person pronoun 'I' and the verb feel. Consequently, SHAP application to the Hybrid Ensemble model helped to understand whether linguistic or psychological factors are taken into account when assessing a person's mental state in this model.

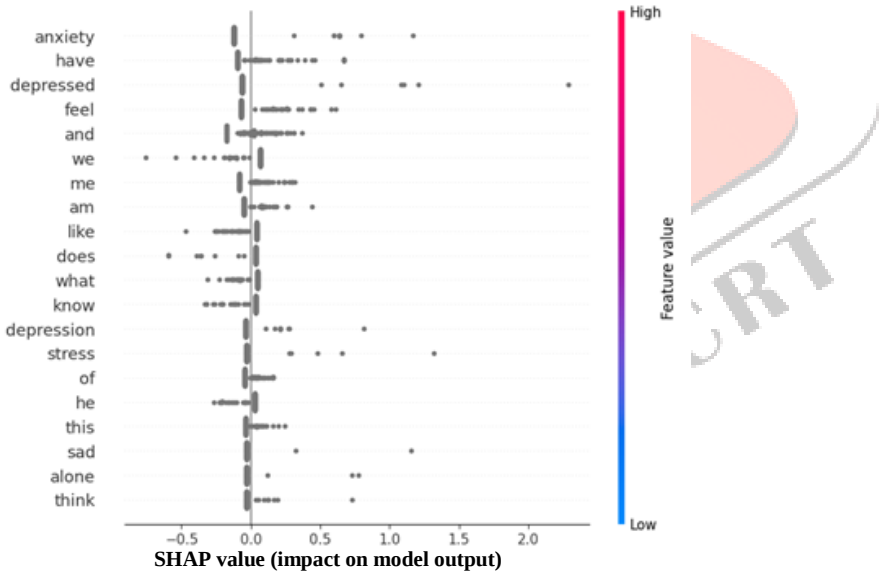


Fig 5: Prediction Explanation Using SHAP-based explainability analysis for Mental Health prediction.

6. 5 COMPARABLE WITH EXISTING STUDIES

The Hybrid model of the ensemble was compared with the results obtained in the previous studies to see how good the model performs. As the previous models are mainly focused on the accuracy of the model, proposed model provides better predictions and at the same time provides a detailed explanation of the result with the help of SHAP values, which makes the result more reliable.

Table 3: Comparable with Existing Studies

Study	Methods	Performance	Explainability	Comparison
Rahman et al. (2020)	SVM, LR, RF	75%	Low	Limited generalization
Iyortsuun et al. (2023)	ML + DL	88%	Moderate	Better accuracy, low explainability
Zafar et al. (2024)	ML, NLP	82%	Low	Lower accuracy, no interpretability
Proposed Work	LSTM + Ensemble	**92. 50%**	**High (SHAP)**	Best performance + explainability

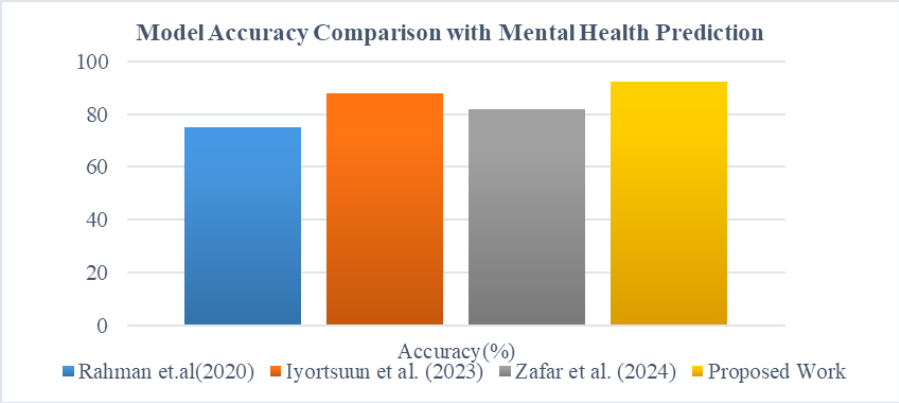


Fig 6: Model Accuracy Comparison with Previous Works

The proposed model outperforms previous works in both accuracy and explainability. So it shows that the proposed model is more reliable.

7. DISCUSSION AND IMPLICATIONS

The results of the experimentation using the mental health prediction system validate that it can indeed be used to ascertain whether or not a person is suffering from a particular mental health condition or not. It can be stated that based on the data provided, that an approach that involves a combination of deep learning, conventional machine learning, and transparency-promoting artificial intelligence can be used to predict mental health risks. The use of a BERT-based LSTM and a Hybrid ensemble method resulted in a considerable improvement in terms of accuracy for the mental health prediction problem, and using a technique for explainability, it is possible to generate interpretable outputs from the model.

7. 1 MODEL PERFORMANCE ANALYSIS

From the results of experiments, it can be seen that the BERT-based LSTM model can successfully capture both the context and sequentiality of text data. As a result of this, BERT-based LSTM provides superior classification performance versus many typical machine learning classifiers. Common classifiers such as Logistic Regression, Random Forest, and Support Vector Machine do not perform well on their own and are not sufficiently performant when compared to the BERT-based LSTM model.

7. 2 HYBRID ENSEMBLE EFFECTIVENESS

It was observed that the prediction strength of the Hybrid ensemble technique was very high. There is sample evidence to prove that the prediction power of this Hybrid ensemble technique is much stronger compared to conventional statistical techniques. The reason for this is that this Hybrid ensemble technique can use all the information that is available with it. The performance of this Hybrid ensemble technique improved manifold for all the parameters, thus proving the use of this ensemble technique for solving this problem.

7. 3 MULTIMODAL INTEGRATION

Incorporating both text and clinical data is the main benefit of this system. For example, clinical items such as PHQ-9 and GAD-7 will help provide a much more accurate assessment of mental health. The use of multiple feature types creates a much better prediction of future outcomes based on these types of features. Additionally, having the clinical override increases the reliability of the system by allowing clinicians to assess potential risks with higher accuracy than without a clinical override.

7. 4 EXPLAINABILITY

The SHAP method can help to provide an understanding of how a model makes its decisions and to help improve transparency and understanding of the model's results in sensitive areas such as mental health, which is essential. The SHAP method will help to identify which features were most significant for each prediction or outcome. An explanation will also help to build user confidence in an AI-based system and to encourage use in health care settings.

7. 5 PRACTICAL IMPLICATIONS

The system is very likely to be used in digital health platforms from a practical point of view. The proposed approach for the predictive modeling using a chatbot is likely to enable the continuous tracking and engagement with the user. This will enable the early identification, which will result in interventions. If the proposed system is successfully implemented, the system will enable the identification of at-risk individuals. In addition, the system will improve the decision-making process. The system is scalable since the system will rely on publicly available data. Therefore, the system is likely to be practical.

8. LIMITATIONS AND FUTURE SCOPE

There are a few disadvantages of the system. The system is mainly based on text data, which will limit the system in terms of generalisation. Additionally, the clinical data such as PHQ-9 and GAD-7 are used only for inference and not for learning. This will limit the system in terms of learning. The use of Text authored by users will increase the level of noise. Additionally, the use of BERT and LSTM has increased the computational complexity of the system. Although the use of SHAP has improved the transparency of the system, it still requires some level of domain knowledge in order to understand the output of the system. Among other developments, there was extensive work done to improve and optimize the addition of several types of multimodal data to the system, including but not limited to speech and behavioral characteristics. Fine-tuning the BERT model with domain-specific data has been achieved as well as providing functionality in real-time and for multiple languages. The remaining updates

impacted the optimization and usability of the system as well as the degree to which the system could clearly articulate or explain its rationale for performing actions taken by it.

9. CONCLUSION

The results of this experiment support the efficiency of the proposed hybrid algorithm for predicting mental illnesses. It can be stated therefore that the accuracy of the designed model was also quite high and achieved 92. 50%. In addition, AUC-ROC metric was at the level of 0. 9759. Although the LSTM algorithm has shown high performance in working with sequence recognition, text prediction, it was found that the use of an ensemble of different models has increased its reliability. Thus, the resulting data demonstrates that the integration of traditional machine learning approaches with neural networks is an optimal method for developing models. Moreover, the implementation of SHAP (SHapley Additive exPlanations) makes the suggested method even more explainable due to gaining more insights about how features affect the output of the model and how the model makes decisions. Such an approach is especially valuable in the case of application in the mental health field as transparency and trust become key criteria for practical use of the solution. This feature allows not only proving the results but also making the suggested tool acceptable. In conclusion, it is worth pointing out that the proposed technique enables an efficient combination of the predictive capacity, stability, and interpretation of the developed artificial intelligence system. Thus, the efficiency of the suggested methodology is proven to be extremely high during the implementation of hybrid intelligent systems to monitor the person's mental state.

REFERENCES

[1] L. Nichols, R. Ryan, C. Connor, M. Birchwood, and T. Marshall, "Derivation of a prediction model for a diagnosis of depression in young adults: A matched case–control study using electronic primary care records," Early Intervention in Psychiatry, vol. 12, no. 3, pp. 444–455, 2018, doi: 10. 1111/eip. 12332.

[2] R. A. Rahman, K. Omar, S. A. M. Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," IEEE Access, vol. 8, pp. 183952–183964, 2020, doi: 10. 1109/ACCESS. 2020. 3029154.

[3] F. Zafar et al. , "The role of artificial intelligence in identifying depression and anxiety: A comprehensive literature review," Cureus, vol. 16, no. 3, p. e56472, 2024, doi: 10. 7759/cureus. 56472.

[4] M. Zakariah and Y. A. Alotaibi, "Unipolar and bipolar depression detection and classification based on actigraphic registration of motor activity using machine learning and uniform manifold approximation and projection methods," Diagnostics, vol. 13, no. 14, p. 2323, 2023, doi: 10. 3390/diagnostics13142323.

[5] N. K. Iyortsuun, S. -H. Kim, M. Jhon, H. -J. Yang, and S. Pant, "A review of machine learning and deep learning approaches on mental health diagnosis," Healthcare, vol. 11, no. 3, p. 285, 2023, doi: 10. 3390/healthcare11030285.

[6] Abd-Alrazaq, A. Rababeh, M. Alajlani, B. M. Bewick, and M. Househ, "Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis," Journal of Medical Internet Research, vol. 22, no. 7, p. e16021, 2020, doi: 10. 2196/16021.

[7] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," JMIR Mental Health, vol. 4, no. 2, p. e19, 2017, doi: 10. 2196/mental. 7785.

[8] G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Science, vol. 6, no. 15, 2017, doi: 10. 1140/epjds/s13688-017-0110-z.

[9] Abd-alrazaq et al. , "The performance of artificial intelligence-driven technologies in diagnosing mental disorders: An umbrella review," npj Digital Medicine, vol. 5, p. 87, 2022, doi: 10. 1038/s41746-022-00631-8.

[10] D. B. Olawade, O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt, "Enhancing mental health with artificial intelligence: Current trends and future prospects," Journal of Medicine, Surgery, and Public Health, vol. 3, p. 100099, 2024, doi: 10. 1016/j. glmedi. 2024. 100099.

[11] M. Alhuwaydi, "Exploring the role of artificial intelligence in mental healthcare: Current trends and future directions – A narrative review for a comprehensive insight," Risk Management and Healthcare Policy, vol. 17, pp. 1339–1348, 2024, doi: 10. 2147/RMHP. S461562.

[12] Vaidyam, A. N. , Wisniewski, H. , Halamka, J. D. , Kashavan, M. S. , & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7), 456-464.

[13] Fulmer, R. , Joerin, A. , Gentile, B. , Lakerink, L. , & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR mental health, 5(4), e9782.

[14] Ku, W. L. , & Min, H. (2024, March). Evaluating machine learning stability in predicting depression and anxiety amidst subjective response errors. In Healthcare (Vol. 12, No. 6). MDPI.

[15] machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Informatics, 10(1), 10.

[16] Bond, R. R. , Mulvenna, M. D. , Potts, C. , O'Neill, S. , Ennis, E. , & Torous, J. (2023). Digital transformation of mental health services. Npj mental health research, 2(1), 13.

[17] Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, Carvalho AF, Keshavan M, Linardon J, Firth J. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry. 2021 Oct;20(3):318-335. doi: 10. 1002/wps. 20883. PMID: 34505369; PMCID: PMC8429349.

[18] Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med. 2022 Apr 8;5(1):46. doi: 10. 1038/s41746-022-00589-7. PMID: 35396451; PMCID: PMC8993841.

[19] Chancellor, S. , & De Choudhury, M. (2020). Methods in predictive techniques for mental health. npj Digital Medicine, 3, 43. https://doi. org/10. 1038/s41746-020-0233-7

[20] De Choudhury, M. , Counts, S. , & Horvitz, E. (2013). Predicting depression via social media. ICWSM. https://doi. org/10. 1609/icwsm. v7i1. 14432

[21] Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. Curr Psychiatry Rep. 2019 Nov 7;21(11):116. doi: 10. 1007/s11920-019-1094-0. PMID: 31701320; PMCID: PMC7274446.

[22] Devlin, J. , Chang, M. W. , Lee, K. , & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers. NAACL-HLT. https://doi. org/10. 48550/arXiv. 1810. 04805

[23] Salas-Zárate R, Alor-Hernández G, Salas-Zárate MDP, Paredes-Valverde MA, Bustos-López M, Sánchez-Cervantes JL. Detecting Depression Signs on Social Media: A Systematic Literature Review. Healthcare (Basel). 2022 Feb 1;10(2):291. doi: 10. 3390/healthcare10020291. PMID: 35206905; PMCID: PMC8871802.

[24] Inkster, B. , Sarda, S. , & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (WYSA) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR mHealth and uHealth, 6, e12106. https://doi. org/10. 2196/12106

[25] Madububambachu U, Ukpebor A, Ihezue U. Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review. Clin Pract Epidemiol Ment Health. 2024 Jul 26;20:e17450179315688. doi: 10. 2174/0117450179315688240607052117. PMID: 39355197; PMCID: PMC11443461.

[26] Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth. 2016 Aug 9;4(3):e96. doi: 10. 2196/mhealth. 6020. PMID: 27507641; PMCID: PMC4995352.

[27] M. Kyrou, I. Kompatsiaris and P. C. Petrantonakis, "Deep Learning Approaches for Stress Detection: A Survey," in IEEE Transactions on Affective Computing, vol. 16, no. 2, pp. 499-517, April-June 2025, doi: 10. 1109/TAFFC. 2024. 3455371.

[28] Naslund, J. A. , Aschbrenner, K. A. , Marsch, L. A. , & Bartels, S. J. (2016). The future of mental health care. Epidemiology and Psychiatric Sciences, 25(2), 113–122. https://doi. org/10. 1017/S2045796015001067

[29] Orabi, A. H. , Buddhitha, P. , Orabi, M. H. , & Inkpen, D. (2018). Deep Learning for Depression Detection of Twitter Users. https://doi. org/10. 18653/v1/W18-0609

[30] Saha K, Torous J, Ernala SK, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med. 2019 Nov 25;9(6):1197-1207. doi: 10. 1093/tbm/ibz028. PMID: 30834942; PMCID: PMC6875652.

[31] Torous, J. , Myrick, K. J. , Rauseo-Ricupero, N. , & Firth, J. (2020). Digital mental health and COVID-19. JMIR Mental Health, 7(3), e18848. https://doi. org/10. 2196/18848