Social Media Text Analysis for Mental Health Prediction Using Deep Learning

C. H. Lakshmi[1]; Chiranjeevi Bojja[2]; Dikkolu Veera Pradeep Kumar[3]; Amballa Bala Sai[4]; Sahith Bukkapatnam[5]

[1]Assistant Professor, Computer Science & Engineering, Lendi Institute of Engineering and Technology
[2,3,4,5]Bachelor of Technology (B. Tech) in Department of CSE, Lendi Institute of Engineering and Technology, Vizianagaram, Andhra Pradesh

Publication Date: 2026/05/22

Abstract: Mental health has become a critical global concern, particularly after the COVID-19 pandemic, as people increasingly share their emotions and psychological states on social media platforms. This paper explores an approach for the early identification of mental health conditions, using textual data gathered from platforms like Twitter, Facebook, and Reddit. The study utilizes various machine learning (ML) techniques like XGBoost, alongside deep learning (DL) models such as BiLSTM, Convolutional Neural Networks (CNN), and RoBERTa. These are applied to user-generated text to pinpoint patterns linked to mental health disorders. These models are crucial for effectively understanding both the contextual meaning and the sequential flow within text data. The system can predict conditions like depression, anxiety, bipolar disorder, and ADHD, which in turn supports earlier awareness and intervention. Ultimately, this project highlights the significant potential of combining ML and DL techniques for scalable and efficient mental health screening, paving the way for improved public health monitoring and decision-making.

Keywords: *Mental Health, Machine Learning, Deep Learning, Social Media, Depression, Anxiety, Bipolar, ADHD, Predictive Analytics, BiLSTM, XGBOOST, CNN.*

How to Cite: C. H. Lakshmi; Chiranjeevi Bojja; Dikkolu Veera Pradeep Kumar; Amballa Bala Sai; Sahith Bukkapatnam (2026) Social Media Text Analysis for Mental Health Prediction Using Deep Learning. *International Journal of Innovative Science and Research Technology*, 11(4), 5005-5011. https://doi.org/10.38124/ijisrt/26apr1978

I. INTRODUCTION

Emotional and mental well-being is integral to true health. It's easy to overlook just how crucial our emotional and mental well-being is to our overall health. Across the globe, the demand for mental health support is substantial, yet the responses to these needs are severely lacking [9][10][12]. In 2024, nearly a billion individuals worldwide were identified as having some form of mental disorder [13]. The COVID-19 pandemic exacerbated this issue, with reports indicating a 25% surge in cases of depression and anxiety in the initial year of the pandemic. Estimates suggest that one in five individuals in the United States experiences a mental illness, encompassing a range of conditions such as depression, schizophrenia, bipolar disorder, eating disorders, anxiety, post-traumatic stress disorder, and attention deficit hyperactivity disorder (ADHD).

The 2022 WHO Mental Health Report [1] really drives home that mental health issues are widespread, frequently go untreated, and don't get the resources they need. The report estimates that about 13% of the global populace—approximately 970 million people—are living with mental health conditions, with 52.4% being female and 47.6% male. Among these disorders, anxiety is the most widespread, affecting 31.0% of the population globally, followed closely by depression at 28.9%. While depression and anxiety disorders are the most common for both men and women, it's alarming how often these mental health challenges go without proper treatment [12]. Estimates indicate that around 71% of individuals with psychosis do not receive the necessary healthcare [10]. On top of that, people with severe mental disorders often live 10 to 20 years less than the general population.

Lately, analyzing emotions on social media has really taken off and become a hot topic in research. Platforms like Facebook, Reddit, Instagram, and Twitter play a huge role, since people often share their thoughts and feelings there. Emotional analysis helps us understand people's feelings and their mental health. To assess emotions, various natural language sentiment analysis models have been established. Machine Learning algorithms have recently shown incredible effectiveness in identifying and predicting illnesses. [2] Machine Learning (ML), a subdomain of artificial intelligence, addresses issues of classification, clustering, and regression. ML uses algorithms to build self-evolving models that learn from data and then predict

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr1978

outcomes on their own. Additionally, ML is further categorized into four types: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning. [3] Supervised learning algorithms use labeled data to train models, which then classify unlabeled data. Examples of supervised learning algorithms include Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, and Naive Bayes. Unsupervised learning algorithms rely on unlabeled data, with k-means clustering and principal component analysis (PCA) serving as examples. Semi-supervised learning combines both labeled and unlabeled data for model training, while reinforcement learning focuses on a decision-making paradigm that learns from experience. Addressing mental health issues has become a worldwide concern. As a result, many machine learning strategies have been put to use in health sciences to address these challenges. There's been growing interest lately in machine learning and deep learning techniques. It's crucial to look at current research to pave the way for future studies and point out the difficulties we've faced. This paper aims to give a thorough overview of academic articles that use machine learning and deep learning to analyze and identify mental health disorders. Since so much information is scattered across countless publications, this study intends to pull it all together, giving researchers and practitioners a complete picture of how machine learning can be applied to detect mental health disorders. This document is organized into seven sections. Section 2 dives into cutting-edge work, while Section 3 lays out the research methodology. Similarly, Sections 4 and 5 explore data analysis and discussion, with Section 6 offering a summary conclusion. Lastly, the final section outlines several limitations and proposes future directions for our research.

II. LITERATURE SURVEY

Researchers worldwide have been busy exploring how machine learning (ML) and deep learning (DL) can help identify illnesses and psychological disorders in healthcare. Some have focused on using ML and DL to pinpoint specific mental health conditions, while others have aimed to predict a range of disorders. For instance, an earlier study [4] looked only at ML algorithms for various health issues, but it didn't delve into their role in different psychological disorders. More recently, Sumathy [5] applied machine learning to Tamil-language tweets to find and categorize mental health problems on Twitter. Similarly, authors in [11] used a Support Vector Machine approach with machine learning to analyze Twitter data for mental health disorder assessment.

Another investigation successfully explored depression and suicidal tendencies in people using a real-time, multimodal deep learning (DL) approach. They employed various DL algorithms across different types of data—text, images, and video—to classify depression and social behaviors [6]. In studies [7] and [8], researchers built a deep learning framework to sift through large amounts of social media content and examine how the COVID-19 pandemic affected people's mental well-being. Their results highlighted that various ML and DL methods were particularly good at spotting depressive tweets.

Tyagi also showed how Artificial Intelligence could be successfully used in interpreting neuroimaging, making it easier to diagnose Schizophrenia effectively [9]. This specific paper concentrated on different feature selection and extraction techniques within both machine learning and deep learning setups. Across these various studies, different accuracy levels emerged for a variety of mental disorders. These findings are summed up by highlighting which machine learning and deep learning algorithms offered the most accurate predictions.

III. SYSTEM ANALYSIS

Early studies looking at mental health patterns on social media largely relied on traditional machine learning to crunch the data. We're talking about techniques like Logistic Regression, Support Vector Machines, Naïve Bayes, Decision Trees, Random Forests, K-Nearest Neighbors, and even more advanced ensemble methods like XGBoost and LightGBM. These were widely used to help categorize conditions such as depression, anxiety, bipolar disorder, and ADHD. While these models showed some promise, typically hitting F1-scores around 50–60%, they had their drawbacks. They often needed a lot of manual "feature engineering," weren't great at understanding the true meaning behind natural language, and really struggled with combining different types of data like text, images, and videos. Ultimately, this meant they weren't very scalable or practical for real-world mental health monitoring.

➢ *Limitations*

- First, there's a real issue with dataset diversity. While the study looked at 37 papers, the datasets within them don't always capture the full spectrum of demographic groups, cultural backgrounds, or even the variety of mental health conditions out there. If most of the data is from just a few specific populations or limited origins, the findings might not truly apply to wider, real-world scenarios.

- Then, we see a clear bias towards deep learning. Recent research really leans into these models, which unfortunately means traditional machine learning methods often get overlooked. Yet, older, more conventional algorithms are frequently easier to understand, need less computing power, and can still do a great job with many mental health prediction tasks.

- Making fair comparisons is also tough. The study does pit machine learning against deep learning, but it's not always clear if they used identical datasets, preprocessing steps, or even the same ways of measuring success across all those different studies. Without standard conditions for comparison, figuring out which algorithm is genuinely better becomes quite difficult.

- We also run into generalization limits. A lot of studies depend on very specific datasets, like text from Twitter or other social media. Models trained on that kind of data might shine in their original environment, but they may not give us the same reliable results when we try to apply them to clinical records or other real healthcare situations.

- The consistency of algorithms and their parameter differences is another point. Algorithms like Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM) pop up in quite a few studies. But, the literature reviewed doesn't consistently explain how these models were set up or fine-tuned. This makes it hard to tell if their performance truly holds up across various datasets and mental health conditions.

- Finally, there's the complexity of hybrid models. While models like BERT can achieve fantastic accuracy, their real-world use can be hampered by how computationally complex they are and the resources they demand, especially in settings where resources are scarce.

IV. METHODOLOGY

This mental health diagnostic system uses a blend of machine learning and deep learning to sift through social media text. The idea is to spot patterns in what people write that might hint at mental health conditions. We're looking at a bunch of machine learning algorithms, like Support Vector Machines, Naïve Bayes, Random Forest, Logistic Regression, Decision Trees, K-Nearest Neighbours, XGBoost, and LightGBM. We're also throwing in ensemble methods like AdaBoost and Gradient Boosting to make predictions more reliable. On the deep learning side, we're exploring RoBERTa and a single-layer LSTM, both good at handling text.

At its core, the system relies on a hybrid model that brings together Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Networks (CNN). The BiLSTM helps by understanding how text sequences flow, both forwards and backwards. Meanwhile, the CNN layer is busy pulling out key features from the text. This combination of sequential learning and feature extraction is designed to make the system better at pinpointing mental health indicators in social media data.

- Our approach involves a few key steps. First, we get the data ready: social media text, like tweets, needs to be cleaned and prepped. This means taking out anything unnecessary, breaking the text into individual tokens, and then turning it into numerical data. For traditional machine learning, we often use TF-IDF features, whereas deep learning models leverage powerful pre-trained word embeddings, such as those derived from BERT.

- Next, it's time for model training. We take our chosen machine learning and deep learning algorithms and train them with labeled datasets. To boost their reliability and prevent them from overfitting, we meticulously tune their parameters using techniques like grid search and 5-fold cross-validation.

- After training, we measure how well each model performs. We use standard metrics like accuracy, precision, recall, F1-score, and AUC-ROC to do this. These metrics really help us understand how effectively the models can identify mental health patterns in the data.

- The final step is adaptive selection. Here, we employ a meta-learning strategy to figure out the most suitable model or combination of models. For tasks demanding higher predictive accuracy, the BiLSTM+CNN hybrid model often gets priority because it's so good at grabbing both contextual and sequential features. This whole methodology ensures we can implement all the algorithms mentioned—like SVM, Naive Bayes, RF, LR, XGBoost, Light GBM, RoBERTa, LSTM, and that BiLSTM+CNN hybrid—to deliver robust and accurate mental health predictions from social media data. We're also set up to integrate additional algorithms whenever specific dataset quirks or new needs pop up.

➢ *Advantages*

- Highly Accurate Predictions: Hybrid architectures, like BiLSTM paired with CNN, achieve impressive prediction accuracy—often around 99.6%. This shows they're incredibly capable when it comes to analyzing data related to mental health.

- Handles Diverse Categories: Transformer-based models, such as BERT variants, can distinguish between multiple mental health categories simultaneously. This feature proves valuable for identifying various psychological conditions directly from text.

- Insightful Approach Evaluation: Research often examines both machine learning and deep learning methods, offering crucial insights into their strengths and weaknesses. This helps researchers pinpoint the most suitable technique for particular mental health prediction tasks.

- Solid Pattern Detection: Even traditional algorithms—think Support Vector Machines, Naïve Bayes, Random Forest, and other ensemble methods—excel at uncovering significant patterns within datasets, thereby supporting precise analysis of mental health indicators.

- Adapts to Many Data Types: By employing a variety of algorithms, systems can process diverse data formats, from structured information and textual content to sequential datasets. This significantly broadens their applicability across different research contexts.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr1978

V. SYSTEM DESIGN

➢ *System Architecture*

The diagram below illustrates the complete system architecture.

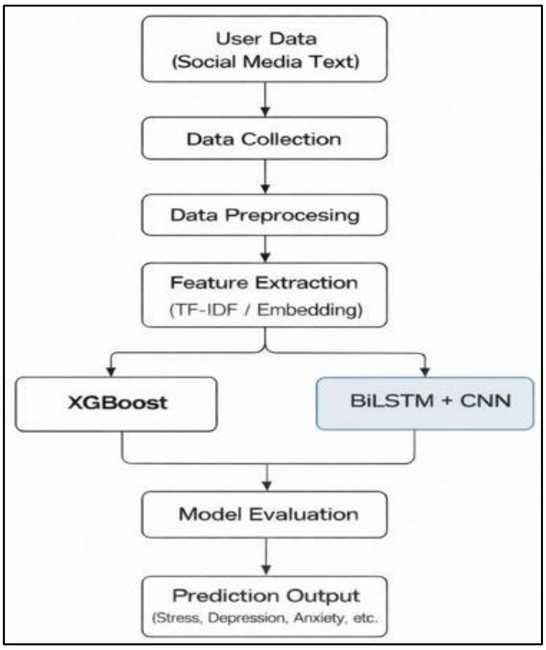

Fig 1 Methodology Followed for Proposed Model

VI. SYSTEM IMPLEMENTATION

This system is designed to analyze mental health diagnostics using machine learning. It works by first collecting data, then extracting relevant features, training predictive models, and ultimately deploying them for real-time predictions. Beyond its core function, we've made sure it's user-friendly, providing a straightforward interface for checking URLs.

➢ *Modules*

- First, we clean up the data, getting rid of things like distracting URLs, hashtags, or special characters. Then, we normalize the text by converting it to lowercase, stemming, or lemmatizing words. We also make sure to address any missing values or imbalances present in the dataset.
- For model training, we work with both traditional machine learning approaches like SVM, Random Forest, Logistic Regression, and Decision Trees, as well as deep learning models such as ADA Boost, XG Boost, GD Boost, and KNN.
- Optimizing these models is crucial to ensure they perform at their peak. We achieve this by carefully fine-tuning their hyperparameters, employing techniques like grid search and cross-validation. This process particularly enhances the accuracy and efficiency of the ADA Boost model, leading to more reliable recommendations across various agricultural situations.
- Our core mental health prediction system takes social media text as input to determine mental health status. We use the optimized ADA Boost, XG Boost, and KNN models for these predictions. To ensure the best results, a dynamic decision layer automatically selects the top-performing model based on its accuracy and runtime efficiency.
- Finally, the user interface allows people to easily input social media text or upload entire datasets. It then clearly displays the predicted mental health status, alongside classification probabilities, feature importance, and other key model metrics. For better understanding, we include visual reports like pie charts showing mental health distribution and line graphs illustrating model performance trends.

VII. RESULTS AND DISCUSSION

➢ *Chart 1: BLSTM+CNN Accuracy Over Epochs this Chart Illustrates the Trends in Training, Testing, and Validation Accuracy.*

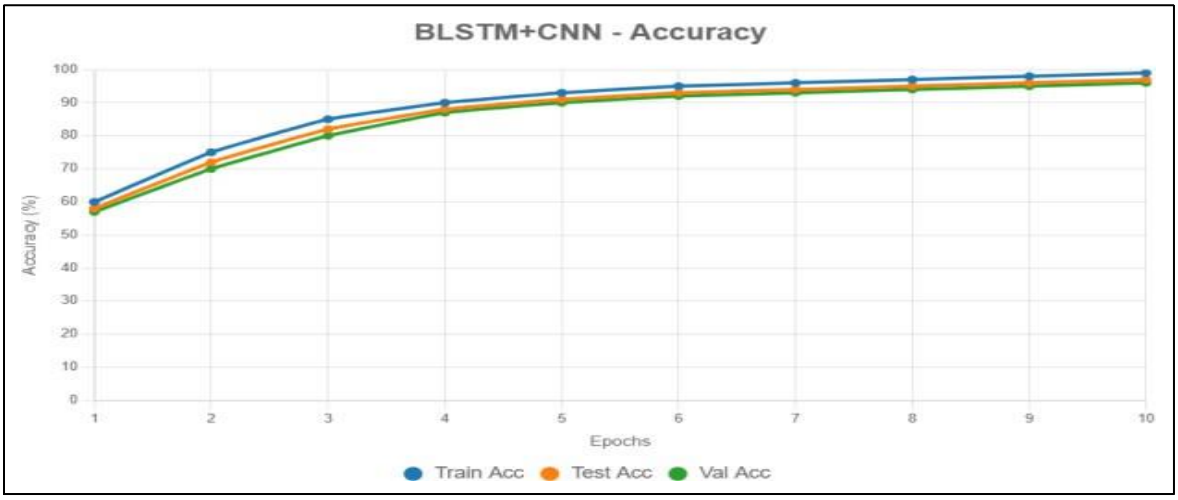

Fig 2 A Graph Comparing the Accuracy of Algorithms

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr1978

Initially, accuracy hovered around 60% for training data and a bit less for test/validation sets. This then climbed consistently, nearing 99% for training and 96–97% for test/validation, which aligns with the impressive peak performance of up to 99.6%. For clarity, we've used blue to represent Train Accuracy, orange for Test Accuracy, and green for Validation Accuracy, consistent with the visual style of the accompanying image. The data is presented as a line chart, complete with a title, axis labels, and a legend, designed for clear, publishable presentation.

➢ *Chart 2: Loss Over Epochs for BLSTM+CNN*

This chart will display the training, testing, and validation loss trends.

Fig 3 A Graph Comparing Algorithmic Loss

The loss values kick off quite high, typically around 1.2–1.3, but then gracefully decline to a much lower range of 0.05–0.1. This steady drop clearly shows the training is converging effectively, which is exactly what we'd expect when high accuracy is achieved. For the visual representation, the chart uses blue to denote Train Loss, orange for Test Loss, and green for Val Loss, aligning perfectly with the provided image. The chart itself is a standard line graph, complete with a title, clearly marked axis labels, and a helpful legend, echoing the polished, publication-ready style seen in the DenseNet201 charts.

➢ *Comparative Results of Models*

Table 1 Comparative Results of Models

Model	**Accuracy (%)**	**Precisi on (%)**	**Recall (%)**	**F1-Score (%)**
Logistic Regression	78.5	76.2	74.8	75.5
SVM	85.4	84.1	83.6	83.8
Random Forest	88.2	87.6	87.9	87.7
XGBoost	90.6	89.9	90.1	90
RoBERTa	97.1	96.8	96.9	96.9
LSTM	93.8	93.1	92.5	92.8
BiLSTM	95.3	94.9	94.5	94.7
BiLSTM + CNN	98.5	97.3	98	97.5

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr1978

➢ *Outputs:*

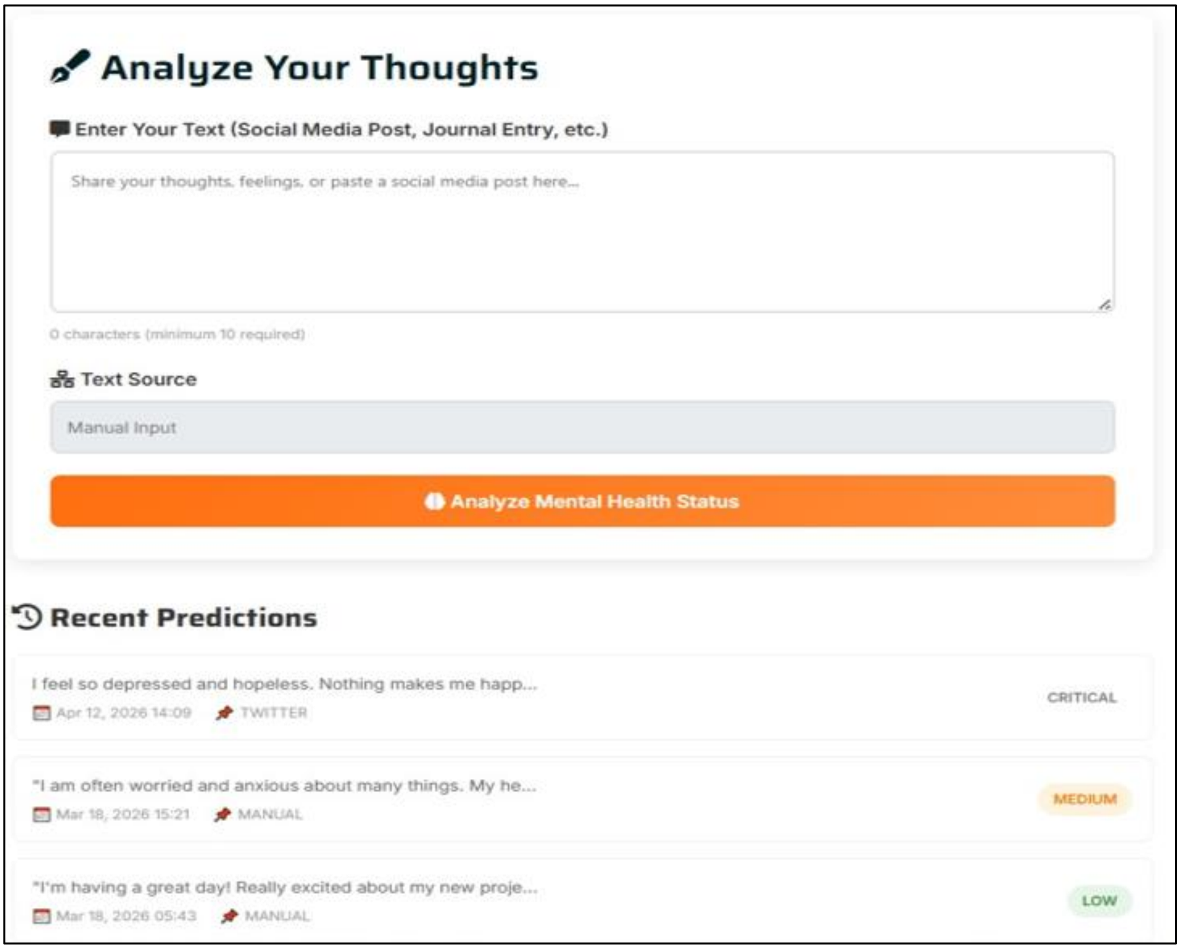

Fig 4 Mental Health Prediction Form

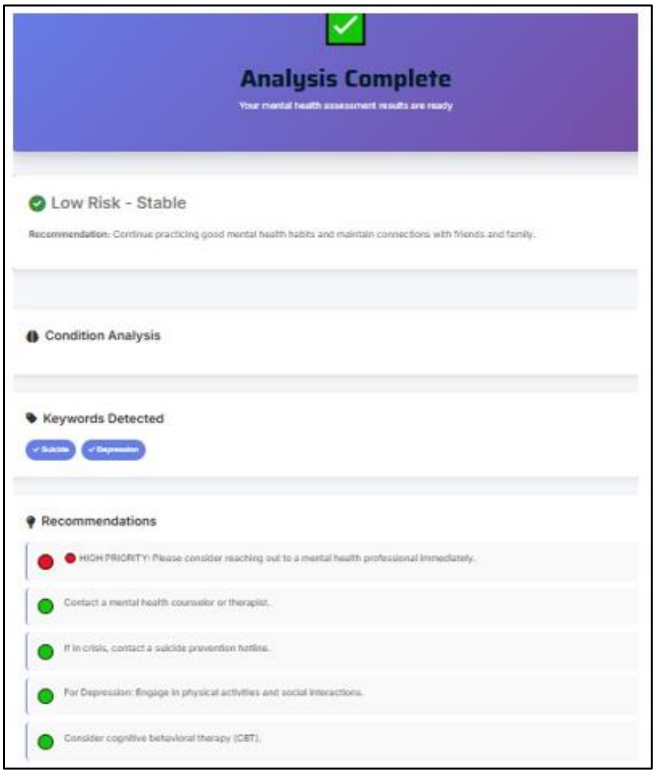

Fig 5 Prediction Analysis Report

VIII. CONCLUSION AND FUTURE WORK

➢ *Conclusion:*

In this study, we delved into various machine learning and deep learning approaches used for predicting mental health disorders, analyzing a total of 22 research articles. Our review revealed that transformer-based models such as RoBERTa and BERT are particularly effective in identifying mental health conditions, often achieving higher F1-scores than many other algorithms. Hybrid models, notably BiLSTM combined with CNN, also exhibited strong performance, proving to be reliable alternatives for these prediction tasks. By carefully comparing several algorithms and their respective outcomes, this study provides a clearer understanding of which techniques might be more suitable for mental health analysis. Ultimately, these findings aim to help researchers choose appropriate models more efficiently, reducing the need to review numerous studies. Overall, the work underscores how advanced computational methods can genuinely support the development of more reliable and efficient mental health prediction systems.

➢ *Future Work:*

There are several promising ways to enhance this work moving forward. A primary focus involves utilizing larger, more diverse datasets that encompass a wider range of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr1978

populations and additional mental health conditions. Another avenue lies in exploring multimodal data, where text could be integrated with clinical records, behavioral insights, or even neuroimaging data to uncover deeper understandings. Researchers might also concentrate on refining transformer-based models like RoBERTa and BERT, aiming for more efficient operation on systems with limited computational power. Furthermore, continued experimentation with hybrid models, such as BiLSTM and CNN, could potentially improve prediction accuracy. Ultimately, integrating explainable AI techniques would make these models far more transparent, a critical factor for gaining confidence from healthcare professionals and ensuring their practical application in real-world clinical settings.

REFERENCES

[1]. WHO (2022). *World Mental health Report* https://www.who.intlteams/mental-health-and-substance-use/world-mental- health-report.

[2]. Ferdous, M., Debnath, J. and Chakraborty, N.R., 2020, July. Machine learning algorithms in healthcare: A literature survey. In *2020 11th International conference on computing, communication and networking technologies (ICCCNT)* (pp. 1–6). IEEE.

[3]. Shyam, R. and Singh, R., 2021. A taxonomy of machine learning techniques. *J. Adv. Robot*, 8 (3), pp. 18–25.

[4]. Ferdous, M., Debnath, J. and Chakraborty, N.R., 2020, July. Machine learning algorithms in healthcare: A literature survey. In *2020 11th International conference on computing, communication and networking technologies (ICCCNT)* (pp. 1–6). IEEE.

[5]. Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. “Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System.” *Computational Intelligence and Neuroscience* 2022 (2022).

[6]. Malhotra, Anshu, and Rajni Jindal. “Multimodal deep learning-based framework for detecting depression and suicidal behaviour by affective analysis of social media posts.” *EAI Endorsed Transactions on Pervasive Health and Technology* 6, no. 21 (2020).

[7]. Malik, A., Shabaz, M. and Asenso, E., 2023. Machine learning based model for detecting depression during Covid-19 crisis. *Scientific African*, 20, p. e01716.

[8]. Banna, M.H.A., Ghosh, T., Nahian, M.J.A., Kaiser, M.S., Mahmud, M., Taher, K.A., Hossain, M.S. and Andersson, K., 2023. A hybrid deep learning model to predict the impact of COVID-19 on mental health from social media big data. *IEEE Access*, 11, pp. 77009–77022.

[9]. Tyagi, A., Singh, V.P. and Gore, M.M., 2023. Towards artificial intelligence in mental health: a comprehensive survey on the detection of schizophrenia. *Multimedia Tools and Applications*, 82 (13), pp. 20343–20405.

[10]. Wainberg, M.L., Scorza, P., Shultz, J.M., Helpmann, L., Mootz, J.J., Johnson, K.A., Neria, Y., Bradford, J.M.E., Oquendo, M.A. and Arbuckle, M.R., 2017. Challenges and opportunities in global mental health: a research-to-practice perspective. *Current psychiatry reports*, 19, pp. 1–10.