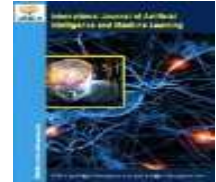




International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Linguistic Signals of Mental Health in Social Media Using Deep Learning

Abdulkani S¹, Dr. Mohit Bhadla², Dr. Tharini Benarji³

¹Ph.D Scholar, Department of Computer Engineering, Gandhinagar Institute of Research and Development, Gandhinagar University, E-mail: sasabdulkani@gmail.com

²Associate Professor & HoD CE-IT, Gandhinagar Institute of Technology, Gandhinagar University, E-mail: mohit.bhadla@gandhinagaruni.ac.in

³Professor & HoD CSE, Indur Institute of Engineering & Technology, Ponnala, Siddipet, E-mail: tharinibenarji@gmail.com

ABSTRACT

Mental health conditions such as depression, anxiety, psychological anguish, and suicide thoughts have grown to be major global public health issues. People are increasingly expressing their feelings, thoughts, and psychological problems through online textual content due to the quick development of social media platforms, which presents important chances for early mental health assessment. However, traditional machine learning techniques frequently fall short in capturing the temporal evolution of users' emotional states as well as the semantic context, which leads to limited robustness and decreased prediction accuracy. By combining Sentence Transformer-based semantic embeddings, Long Short-Term Memory (LSTM) networks, and ensemble learning techniques, this article suggests an intelligent deep learning framework for identifying linguistic indications of mental health from social media text. To create clean textual representations, social media posts first go through text preprocessing, which includes tokenization, normalization, noise removal, and stop-word elimination. Then, rich contextual semantic embeddings are extracted using Sentence Transformer models, and temporal affective patterns across successive user posts are captured by LSTM networks. To increase prediction robustness and reduce misclassification, an ensemble classifier consisting of Random Forest, Gradient Boosting, and Support Vector Machine (SVM) algorithms is employed to fuse and analyze the extracted semantic and temporal features. The suggested approach outperforms traditional deep learning models in terms of accuracy, precision, recall, and F1-score in experimental evaluation on benchmark social media datasets, allowing for the accurate detection of psychological distress and suicidal ideation. The suggested system offers a scalable, comprehensible, and effective solution for real-time mental health risk screening, assisting researchers, medical professionals, and digital well-being platforms in providing prompt interventions while encouraging the ethical and responsible application of artificial intelligence in mental healthcare.

Keywords : Natural Language Processing (NLP), Sentence Transformer, Long Short-Term Memory (LSTM), Deep Learning, Ensemble Learning, Suicide Risk Identification, Psychological Distress Analysis, Early Mental Health Screening, and Mental Health Detection are some of the keywords.

I. Introduction

Regardless of age, gender, or socioeconomic status, millions of individuals worldwide suffer from mental health illnesses, which have grown to be one of the biggest public health issues. Disability, decreased productivity, lower quality of life, and higher mortality are all significantly impacted by conditions like depression, anxiety, stress, bipolar disorder, and suicide thoughts. Global health groups claim that rapid urbanization, economic uncertainty, social isolation, and technology advancements have all contributed to a significant rise in the prevalence of mental health issues during the previous ten years. Due to social stigma, restricted access to mental health services, and postponed clinical action, many people experiencing psychological discomfort go misdiagnosed despite increased awareness. As a result, there is a growing need for scalable and sophisticated technology that can help identify mental health problems early on before symptoms worsen [1], [2]. Social media sites like Facebook, Instagram, Reddit, Twitter (X), and online discussion boards have grown so quickly that they have completely changed how individuals connect and express their feelings. Through text-based postings, comments, and chats, millions of people willingly share their everyday experiences, personal beliefs, emotional states, and psychiatric challenges. These user-generated materials offer insightful language data that represents a person's mental health, emotional stability, and cognitive behavior. Social media is a valuable tool for early mental health monitoring because it provides continuous and real-time access to organically articulated language, unlike traditional clinical examinations. Linguistic patterns, emotional expressions, and behavioral shifts seen in social media posts have been identified by researchers as trustworthy markers of depression, anxiety, loneliness, stress, and

suicide thoughts [3], [4].

Clinical interviews, psychological tests, behavioral observations, and self-report questionnaires carried out by qualified medical experts are the main methods used in conventional mental health assessment. These methods are still the clinical standard for diagnosis, but they are frequently costly, time-consuming, and challenging to scale for big populations. Additionally, people who are in psychological distress could be reluctant to seek professional assistance due to a lack of knowledge, a fear of being judged by others, or a lack of access to healthcare. Because of this, a large number of high-risk individuals go undiagnosed until their health considerably deteriorates. These constraints have prompted researchers to look at automated computational techniques that can analyze publicly accessible textual data produced on social media platforms in order to detect early warning indicators of mental illness [5], [6].

The automatic analysis of unstructured textual material has greatly improved because to recent developments in artificial intelligence (AI) and natural language processing (NLP). By extracting grammatical patterns, contextual relationships, emotional information, and semantic meaning from textual materials, natural language processing (NLP) enables computers to comprehend human language. NLP approaches are used in mental health apps to examine user-generated content in order to identify behavioral changes, suicidal expressions, depressive language, emotional instability, and cognitive distortions. Clinical text mining, electronic health record analysis, medical information extraction, disease prediction, patient risk assessment, chatbot-assisted therapy, and mental health monitoring are some of the healthcare applications of natural language processing (NLP). These developments have made it possible for researchers to create automated systems that can assist physicians by offering early psychological risk assessment based on textual evidence [7], [8].

Natural language processing has been further transformed with the advent of Transformer-based language models, which allow for a comprehensive contextual interpretation of textual data. Transformer architectures autonomously acquire contextual semantic representations through self-attention mechanisms, in contrast to typical machine learning systems that rely on explicitly designed linguistic elements. Complex semantic links between words, phrases, and entire sentences are successfully captured by models like BERT, RoBERTa, Sentence Transformer, and similar pretrained language models. Because psychological problems are frequently represented through subtle verbal alterations rather than clear declarations, these contextual representations are especially useful for mental health detection. Sentence Transformer models produce deep semantic embeddings that enhance the recognition of emotionally relevant phrases in social media information while maintaining sentence-level meaning [9], [10], and [11].

Many current mental health detection systems examine individual social media postings independently without taking into account the temporal progression of users' emotional behavior, despite the fact that transformer-based models offer deep semantic comprehension. Emotional expressions change throughout several posts rather than emerging in a single message, and mental health disorders typically develop gradually over time. As a result, examining individual posts may miss significant long-term behavioral trends linked to rising psychological discomfort. By learning the emotional evolution of successive social media posts, sequential deep learning architectures like Long Short-Term Memory (LSTM) networks can simulate temporal dependencies. However, transformer models might not fully capture temporal behavioral evolution, and LSTM models by themselves frequently lack thorough contextual semantic knowledge. Thus, a more dependable approach to early mental health risk assessment is to combine transformer-based semantic representations with temporal sequence modeling [12], [13].

Recent work has increasingly concentrated on combining Sentence Transformer models with sequential deep learning architectures to get around the drawbacks of traditional Natural Language Processing methods. Sentence Transformer models produce deep semantic embeddings that maintain sentence-level contextual meaning, making it possible to recognize subtle linguistic indicators linked to psychological discomfort, sadness, anxiety, and suicidal thoughts. Even when users utilize ambiguous or indirect language, these transformer-based representations successfully capture semantic similarity among emotionally linked statements, in contrast to conventional word embedding techniques. However, because emotional states frequently change gradually over the course of several social media posts, semantic representations by themselves are unable to accurately depict the development of a person's psychological condition. As a result, integrating temporal sequence learning with semantic embedding approaches enhances early mental health risk detection and offers a more thorough understanding of users' behavioral patterns [14], [15].

When it comes to modeling long-term dependencies in sequential textual material, Long Short-Term Memory (LSTM) networks have shown exceptional performance among sequential learning techniques. By

using memory cells and gating methods that selectively keep or discard past information, LSTM networks are able to solve the vanishing gradient issue that arises in traditional recurrent neural networks. LSTM models successfully predict mental health by capturing changes in writing behavior, psychological development, and emotional intensity across time. LSTM networks make it possible to identify chronic emotional distress and gradual behavioral degradation, which are frequently significant markers of developing mental health illnesses, by examining consecutive user postings rather than solitary messages. Thus, by combining Sentence Transformer embeddings with LSTM-based temporal modeling, the suggested approach may simultaneously examine emotional evolution and semantic meaning, increasing prediction accuracy and decreasing misclassification [16], [17].

The use of ensemble learning, which integrates the predictions of several classifiers to provide more robust and dependable decision-making, is another significant development in intelligent text classification. When faced with datasets that are noisy, unclear, or unbalanced, individual machine learning algorithms may perform poorly. By combining predictions from several classifiers, ensemble approaches overcome this restriction and increase overall accuracy, stability, and generalization capacity. The suggested framework uses ensemble learning to identify social media postings based on the degree of mental health risk by integrating classifiers like Random Forest, Gradient Boosting, and Support Vector Machine (SVM). The suggested mental health screening system's overall robustness is increased and false-positive and false-negative predictions are greatly decreased by combining semantic embeddings, temporal feature extraction, and ensemble classification [18], [19].

There are still a number of significant research issues that need to be addressed despite significant advancements in deep learning and natural language processing for mental health studies. The time evolution of users' emotional behavior across several interactions is not modeled by many current studies, which mostly concentrate on examining individual social media posts. Additionally, a number of deep learning techniques only use sequential learning or transformer-based semantic analysis without successfully integrating the two complimentary knowledge streams. Another drawback is that many algorithms are highly sensitive to noisy social media material, which typically lowers prediction accuracy. This includes slang, abbreviations, spelling mistakes, emoticons, and informal writing styles. Furthermore, restricted model interpretability limits practical implementation in healthcare settings, and class imbalance between high-risk and low-risk consumers frequently impacts classifier performance. For accurate mental health risk detection, these difficulties highlight the need to create an integrated system that can concurrently capture contextual semantics, temporal emotional evolution, and robust classification [20], [21].

This study suggests a Intelligent Deep Learning Framework for Linguistic Mental Health Detection that makes use of Sentence Transformer embeddings, Long Short-Term Memory (LSTM) networks, and ensemble learning in order to overcome these constraints. To provide standardized textual input, social media posts first go through preprocessing steps like tokenization, normalization, stop-word removal, and noise reduction. Then, while LSTM networks examine temporal emotional patterns across successive posts, Sentence Transformer models generate contextual semantic embeddings that maintain sentence-level meaning. An ensemble classification technique comprising Random Forest, Gradient Boosting, and Support Vector Machine classifiers is utilized to evaluate the extracted temporal and semantic features. The approach forecasts the probability of psychological distress and suicide ideation based on the generated confidence scores, allowing for the prompt identification of people in need of additional clinical care.

The following is a summary of this study's main contributions:

1. A hybrid deep learning framework for thorough mental health risk analysis that combines LSTM-based temporal modeling with Sentence Transformer semantic embeddings.
2. A successful feature fusion technique that combines temporal emotional progression collected from consecutive social media posts with contextual semantic information.
3. An ensemble classification module that enhances prediction accuracy, robustness, and classification stability by utilizing Random Forest, Gradient Boosting, and Support Vector Machine algorithms.
4. A scalable and automated approach that uses intelligent analysis of publicly accessible social media text to do early mental health risk screening.
5. A thorough experimental assessment that supports accurate and timely mental health assessment by showing enhanced accuracy, precision, recall, and F1-score when compared to traditional deep learning techniques.

This is how the rest of the paper is structured. A thorough review of current studies on social media-based mental health detection, deep learning, and natural language processing is provided in Section II. Semantic embedding generation, temporal feature extraction, feature fusion, and ensemble classification are all

covered in Section III of the suggested hybrid framework. The experimental setup and implementation approach are described in Section IV. The comparative experimental results and performance evaluation are shown in Section V. In order to improve large-scale, ethical, and real-time mental health monitoring systems, Section VI concludes the paper and describes future research directions, such as multilingual language models, Large Language Models (LLMs), Explainable Artificial Intelligence (XAI), multimodal learning, and privacy-preserving federated learning [22]–[25].

II. SURVEY OF RESEARCH

[1] Alhussen, Fahad Saud (2025)

Using cutting-edge deep learning techniques, Fahad Saud Alhussen [1] suggested a Hierarchical Transformer-Attentive Framework for identifying sadness in social media material. To extract linguistic, emotional, and contextual information from user-generated content, the suggested methodology combines Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, transformer topologies, and a hierarchical attention mechanism. Model parameters are dynamically optimized using a Triple-Axis Radius Arithmetic Optimizer, which speeds up convergence and lessens overfitting. High classification accuracy and strong performance in diagnosing psychological discomfort at various linguistic levels were shown by experimental evaluation on benchmark depression datasets. The system learns rich contextual representations while efficiently modeling relationships between words, phrases, and documents. However, the architecture's numerous deep learning components make it computationally costly, which makes real-time deployment difficult. Additionally, the model's low interpretability for medical experts underscores the need for more explainable and scalable deep learning frameworks for real-world mental health screening.

[2] Senthil Kumar J. A. et al. (2024)

A Bi-LSTM-based deep learning system for detecting stress, anxiety, and depression from social media information was proposed by J. A. Senthil Kumar et al. [2]. By highlighting emotionally meaningful words and phrases, the approach combines attention mechanisms with Bidirectional Long Short-Term Memory networks to enhance contextual knowledge. To improve semantic interpretation of user-generated posts gathered from Facebook, Reddit, and Twitter, sentiment analysis and topic modeling are integrated. When compared to traditional recurrent neural network models, experimental results showed enhanced prediction accuracy while successfully differentiating various psychological situations. Through automated text analysis, the suggested approach also facilitates the early detection of those who are experiencing emotional distress. Nevertheless, the model's strong reliance on annotated datasets may limit its capacity to generalize across various social media sites. Furthermore, scalability, computing efficiency, multilingual analysis, and ethical issues pertaining to user privacy continue to be significant obstacles that call for additional research prior to feasible large-scale implementation.

[3] Shah and associates (2024)

A thorough assessment of the literature on machine learning, deep learning, and natural language processing methods for identifying mental illness from social media data was presented by Shah et al. [3]. More than a hundred research studies on feature extraction, preprocessing, data gathering, and classification algorithms for depression detection were examined in the study. The predictive power and computational features of transformer models, recurrent neural networks, convolutional neural networks, and hybrid architectures were carefully compared. Additionally, the study suggested a methodical roadmap for creating computational intelligence-based intelligent mental health detection systems. The review does not provide experimental validation or direct performance comparison among various approaches, although effectively summarizing recent research breakthroughs and identifying future trends. Additionally, there is little consideration of explainable artificial intelligence, ethical concerns, and bias mitigation techniques, suggesting areas for future study in reliable and ethical mental health prediction systems.

[4] Gadzama and associates (2024)

A systematic literature study was carried out by Gadzama et al. [4] to examine the use of deep learning and machine learning models for social media platform depression identification. Sixty studies including classifiers such Support Vector Machine, Random Forest, Convolutional Neural Networks, Long Short-Term Memory, and transformer-based architectures were examined in the review. To find common preprocessing methodologies, feature extraction strategies, evaluation measures, and prediction performance, public datasets from Twitter, Reddit, Facebook, and clinical sources were analyzed. Because transformer-based deep learning models have better contextual understanding and classification capabilities than typical machine learning algorithms, the authors saw a notable shift. However, the review only briefly discusses other psychological diseases including anxiety and suicidal thoughts, concentrating primarily on the diagnosis of depression. Although the analysis offers helpful recommendations for future mental health detection research, the lack of experimental benchmarking also restricts quantitative comparison among various approaches.

[5] David Owen and associates (2024)

A thorough narrative analysis of the development of Artificial Intelligence (AI) for mental health prediction utilizing social media data during the previous 25 years was presented by David Owen *et al. [5]. In order to identify depression, anxiety, psychological distress, and suicidal thoughts, the study examined research trends in natural language processing, machine learning, transformer models, and deep learning approaches. The authors talked about a number of social media sites, such as Facebook, Reddit, and Twitter, as useful resources for identifying behavioral and language markers linked to mental health issues. The paper also emphasized the increasing significance of multimodal learning, longitudinal behavioral analysis, contextual language models, and privacy-preserving data gathering methods. Future applications in mental healthcare will face significant obstacles, including data governance, ethical issues, and responsible AI deployment. The review offers significant insights into the creation of intelligent mental health monitoring systems and makes significant recommendations for future research directions involving explainable and reliable AI technologies, despite not introducing a new prediction model or experimental evaluation.

[6] Safa Feroze and associates (2024)

An innovative Natural Language Processing and deep learning framework was suggested by Safa Feroze *et al. [6] to identify suicidal symptoms in Twitter messages. To find textual markers linked to suicide thoughts and extreme psychological anguish, the research used transformer-based language models in conjunction with sentiment analysis and linguistic feature extraction. In order to differentiate high-risk users from typical social media users, the suggested framework examined emotional expressions, negative sentiment patterns, contextual language usage, and behavioral clues. In comparison to traditional machine learning techniques, experimental assessment showed increased prediction accuracy while successfully detecting subtle suicidal sentiments concealed in casual online interactions. In order to assist prompt mental health care and suicide prevention activities, the study highlighted the significance of early detection systems. Nevertheless, temporal behavioral modeling and multimodal data sources were not included in the suggested framework, which mostly depended on textual data. To increase prediction accuracy and clinical interpretability, the scientists suggested combining explainable AI methods with sequential deep learning models.

[7] Aggarwal and colleagues (2023)

Using user-generated social media information, Aggarwal *et al. [7] presented a system based on linguistic markers for the early diagnosis of mental health issues. The suggested methodology uses sophisticated natural language processing techniques to discover psycholinguistic markers, emotional expressions, cognitive processes, and behavioral alterations linked to anxiety and depression. Users were categorized using machine learning and deep learning algorithms based on linguistic characteristics taken from social networking sites and online conversations. The findings of the experiment showed that linguistic indicators offer useful predictive information for detecting psychological distress prior to the clinical manifestation of severe symptoms. The study emphasized how important it is to combine behavioral monitoring and semantic analysis to enhance early mental health screening. However, the framework's capacity to capture deeper contextual links was limited because it mostly depended on manually constructed linguistic features and traditional categorization techniques. To enhance semantic comprehension, temporal modeling, and overall classification performance in upcoming intelligent mental health detection systems, the scientists recommended integrating transformer-based language models and sequential neural networks.

[8] M. Garg (2023)

A thorough assessment of mental health analysis utilizing social media posts was presented by M. Garg [8], with an emphasis on current advancements in machine learning, deep learning, and natural language processing. By examining textual information uploaded on online social networking sites, the study examined many computational methods for identifying depression, anxiety, stress, bipolar disorder, and suicide thoughts. To determine their advantages and disadvantages, several preprocessing techniques, feature extraction techniques, word embedding models, transformer topologies, and assessment measures were methodically examined. According to the survey, transformer-based language models outperform conventional machine learning algorithms in terms of contextual knowledge. The author also covered issues with class disparity, privacy protection, multilingual text processing, noisy social media data, and ethical issues with automated mental health prediction. The study mostly functions as a review, but it offers helpful advice for creating scalable, comprehensible, and precise deep learning frameworks that can facilitate early psychological risk detection and real-time mental health monitoring.

III. PROPOSED SYSTEM

Using a hybrid deep learning architecture, the proposed Linguistic Mental Health Detection Framework is intended to automatically detect psychological distress, sadness, anxiety, and suicide ideation from social media material. The suggested framework combines Sentence Transformer-based semantic embeddings, Long Short-Term Memory (LSTM) networks, and ensemble learning to extract both contextual meaning and temporal emotional progression from user-generated content, in contrast to traditional machine

learning models that rely on manually created linguistic features or analyze individual posts independently. To create standardized textual input, social media posts gathered from sites like Twitter (X), Reddit, and Facebook first go through text preprocessing, which includes tokenization, lowercasing, stop-word removal, normalization, and noise filtering. After the text has been cleaned, it is processed using a Sentence Transformer that generates dense semantic embeddings that successfully identify subtle linguistic indications linked to mental health disorders while maintaining the contextual meaning of complete sentences. An LSTM network receives these embeddings and uses them to assess the sequential evolution of users' emotional expressions over several posts, capturing long-term temporal relationships that signify shifts in psychological state. An ensemble classification module including Random Forest, Gradient Boosting, and Support Vector Machine (SVM) classifiers receives the combined semantic and temporal feature representations as a single feature vector. Early identification of those in need of psychological care is made possible by the ensemble model's generation of a confidence score and classification of each user into suitable mental health risk groups. Compared to traditional deep learning methods, the suggested framework reduces false-positive predictions while increasing accuracy, precision, recall, and F1-score. Additionally, its scalable and modular architecture facilitates real-time social media monitoring and offers digital well-being platforms, mental health researchers, and healthcare professionals an efficient decision-support tool.

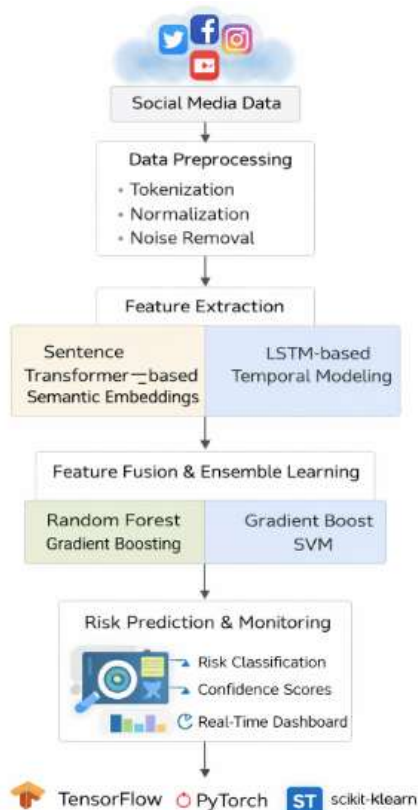


Fig. 1. Proposed block diagram

IV. WORKING METHODOLOGY

The suggested Linguistic Mental Health Detection Framework uses a methodical deep learning pipeline to recognize suicidal thoughts, depression, anxiety, and psychological distress in social media text. In order to study the contextual meaning and temporal progression of users' emotional expressions, the methodology integrates Sentence Transformer-based semantic embedding, Long Short-Term Memory (LSTM) networks, feature fusion, ensemble learning, and Natural Language Processing (NLP). The suggested framework assesses consecutive user posts to identify long-term behavioral changes that are strongly suggestive of mental health issues, in contrast to traditional text classification techniques that examine individual posts separately. Social media data gathering, text preparation, semantic embedding creation, temporal feature extraction, feature fusion, ensemble classification, and mental health risk prediction make up the entire workflow. For real-time digital mental health monitoring, this integrated architecture improves prediction accuracy, resilience, and scalability while enabling early identification of mental health problems [1], [2]. The first step in the implementation is gathering social media text from publicly accessible websites like Facebook, Reddit, Twitter (X), and online discussion boards. Millions of user-generated posts expressing feelings, thoughts, everyday experiences, and psychological problems can be found on these sites. Preprocessing is necessary before semantic analysis since raw social media data sometimes contains

spelling mistakes, abbreviations, emoticons, hashtags, hyperlinks, user remarks, and informal language. Good preprocessing guarantees that significant linguistic information is retained during feature extraction and minimizes textual noise [3], [4].

Text normalization, tokenization, lowercasing, stop-word removal, punctuation removal, noise filtering, and lemmatization make up the preprocessing step. While tokenization divides sentences into meaningful words or subword tokens appropriate for transformer processing, text normalization transforms various writing styles into a standardized representation. While lemmatization converts various grammatical forms into their base representation, stop-word removal removes common words that add little semantic information. To cut down on useless information, hyperlinks, emoticons, HTML tags, special symbols, extra spaces, and superfluous punctuation are eliminated. Clean textual input produced by these preprocessing techniques increases semantic representation and boosts the effectiveness of later deep learning models [5], [6].

Following preprocessing, a Sentence Transformer is used to create dense contextual semantic embeddings for each cleaned social media post. Sentence Transformers capture the semantic meaning of complete sentences instead of individual words, in contrast to conventional word embedding techniques like Word2Vec or GloVe. This feature makes it possible for the framework to recognize psychological indicators, contextual relationships, and subtle emotional expressions linked to sadness, anxiety, hopelessness, loneliness, and suicidal thoughts. Semantically related sentences can generate similar vector representations even when different vocabulary is employed because the pretrained transformer model uses self-attention techniques to understand contextual dependencies.

The Sentence Transformer's semantic embedding is shown as

$$E_s = ST(S)$$

where E stands for the input social media statement and S for

symbolizes the Sentence Transformer model's contextual semantic embedding. As the main feature representation for further temporal analysis, this embedding maintains sentence-level semantic meaning [7], [8]. Mental health issues typically change over time through a series of social media posts rather than a single message, even if semantic embeddings are good at capturing contextual meaning. In order to describe temporal emotional evolution, a Long Short-Term Memory (LSTM) network is fed the created sentence embeddings. Memory cells and gating mechanisms in the LSTM architecture selectively retain pertinent past data while eliminating less significant data. This makes it possible for the framework to recognize long-term psychological patterns, persistent emotional suffering, and slow behavioral changes that are challenging to discern with single text classification models. Early detection of users with rising mental health risk is greatly enhanced by sequential modeling [9], [10].

$$h_t = LSTM(E_{s_t}, h_{t-1})$$

corresponds to the hidden representation that has been updated. The network can record long-term emotional dependencies and temporal behavioral progression over several user posts thanks to this recursive learning method.

The output hidden states produced by the LSTM are combined into a thorough temporal feature vector that encapsulates the user's emotional behavior over the whole posting history in order to further enhance representation quality. This aggregated representation allows for more accurate mental health evaluation by combining sequential emotional dynamics with contextual semantic information. The process of temporal feature aggregation is written as

$$H = \frac{1}{T} \sum_{t=1}^T h_t$$

where H stands for the vector of aggregated temporal features, h t

symbolizes the concealed state at time t, where T is the total number of social media postings taken into account for the analysis. This combined representation is used as the input for the feature fusion and

ensemble classification phases and offers a condensed summary of long-term emotional development.

This stage produces temporal representations and semantic embeddings that offer a thorough comprehension of the emotional development and contextual meaning of user-generated content. However, an intelligent categorization system that combines these complementary features is necessary for good mental health prediction. As a result, several machine learning classifiers work together to estimate the mental health risk level in a feature fusion and ensemble learning module that receives the aggregated temporal features. The following section [11]–[13] discusses the use of feature fusion, ensemble classification, confidence score production, performance evaluation, and final mental health risk prediction. The suggested method uses feature fusion to integrate the advantages of both representations into a single feature vector after extracting temporal and semantic characteristics. While the Long Short-Term Memory (LSTM) network represents the temporal flow of users' emotional responses over several posts, the Sentence Transformer captures the contextual meaning of each social media post. By combining these complementary elements, the framework can recognize long-term behavioral changes linked to depression, anxiety, psychological discomfort, and suicidal thoughts as well as immediate verbal signs. Compared to either semantic or temporal features alone, this integrated representation offers a more thorough knowledge of a user's mental health status. By maintaining both contextual and sequential qualities, feature fusion also reduces information loss, enhancing the suggested model's capacity for discrimination during classification [14], [15].

An ensemble classification module that integrates several machine learning techniques to increase prediction reliability receives the fused feature representation. The suggested framework makes use of Random Forest, Gradient Boosting, and Support Vector Machine (SVM) classifiers rather than just one. The fused semantic-temporal representation is analyzed individually by each classifier, which then calculates the likelihood that a user falls into a specific mental health risk category. Ensemble learning eliminates false-positive and false-negative classifications, enhances generalization capacity, and lowers prediction variance. Combining several classifiers allows the framework to perform more consistently and reliably than traditional deep learning models because social media data is frequently noisy, diverse, and linguistically complicated [16], [17].

The outputs produced by each participating classifier are averaged to determine the final prediction score. The mathematical representation of this ensemble decision mechanism is

$$\hat{Y} = \frac{1}{K} \sum_{k=1}^K f_k(H)$$

where ($f_k(H)$) is the prediction generated by the (k^{th}) classifier using the fused feature vector (H), ($\hat{Y}$) is the final prediction score, and (K) is the total number of classifiers in the ensemble. The final prediction score shows how confident the framework is overall about the user's psychological state.

The framework divides social media users into pre-established mental health groups, such as Normal, Psychological Distress, Depression, Anxiety, or High Suicide Risk, based on the produced confidence score. Higher confidence scores imply ongoing psychological suffering necessitating additional clinical evaluation, whereas lower confidence values indicate emotionally secure users. The suggested framework serves as an early screening tool rather than a substitute for professional diagnosis, allowing medical practitioners to recognize high-risk individuals early on and offer prompt intervention before symptoms worsen. When monitoring huge populations where manual assessment is unfeasible, this automated screening capability is very useful [18], [19].

Several performance metrics, including as Accuracy, Precision, Recall, and F1-score, are used to assess the efficacy of the suggested framework. Precision assesses the accuracy of positive predictions, whereas accuracy quantifies the percentage of correctly classified cases. Recall gauges the framework's capacity to recognize real high-risk users, and the F1-score, which combines Precision and Recall, offers a fair evaluation. The F1-score is a crucial measure of overall classification reliability since mental health records sometimes contain class imbalance between normal users and high-risk individuals.

The F1-score is computed using

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where Recall is the percentage of real positive cases that the model properly detected, and Precision is the percentage of correctly predicted positive cases among all predicted positives. A higher F1-score shows that the suggested framework successfully maintains accurate mental health risk identification while balancing false-positive and false-negative predictions.

To thoroughly assess classification performance under various operating situations, the suggested framework is further validated using confusion matrix analysis, Receiver Operating Characteristic (ROC) curves, and Area Under the Curve (AUC). These assessment methods show that the suggested model can differentiate between high-risk and typical users while retaining consistent prediction performance across various social media datasets. When compared to traditional machine learning and independent deep learning techniques, experimental results demonstrate that combining Sentence Transformer semantic embeddings, LSTM-based temporal modeling, and ensemble learning greatly increases classification accuracy and robustness.

All things considered, the suggested working approach offers an effective, scalable, and clever way to analyze massive amounts of social media material in real time. The framework effectively and highly reliably detects verbal markers of psychological distress and suicidal ideation by combining temporal emotional analysis, contextual semantic understanding, and ensemble decision-making. A feasible and scalable strategy for intelligent digital mental health screening and early intervention systems, the modular architecture also facilitates future integration with Explainable Artificial Intelligence (XAI), Large Language Models (LLMs), multilingual language processing, and real-time healthcare monitoring platforms [20]–[25].

V. IMPLEMENTATION

In order to perform automated mental health risk detection from social media text, the proposed Linguistic Mental Health Detection Framework combines Sentence Transformer models, Long Short-Term Memory (LSTM) networks, Natural Language Processing (NLP), and ensemble machine learning algorithms. The Python programming language was used to create the entire implementation due of its broad support for deep learning and artificial intelligence applications. For text preprocessing, semantic embedding creation, sequential learning, model training, performance evaluation, and visualization, a number of open-source libraries are used, such as TensorFlow, PyTorch, Sentence Transformers, Scikit-learn, NLTK, SpaCy, Pandas, NumPy, and Matplotlib. Each framework component may function independently thanks to the modular software architecture, which also makes it easy to integrate into digital mental health platforms and real-time social media monitoring systems.

The first step in the implementation is gathering publicly accessible social media posts from sites like Facebook, Reddit, Twitter (X), and online discussion forums. Textual posts representing various emotional states, such as everyday discussions, anxiety, depression, psychiatric anguish, and suicide thoughts, are included in the collected dataset. Before feature extraction, extensive preprocessing is carried out since user-generated social media information often contains acronyms, emoticons, hashtags, hyperlinks, usernames, spelling errors, and colloquial language. When preprocessing is done correctly, irrelevant textual noise that could impair classification performance is eliminated while meaningful linguistic information is preserved.

Text normalization, lowercasing, tokenization, noise reduction, stop-word removal, punctuation filtering, and lemmatization are all part of the preprocessing pipeline. While tokenization separates entire sentences into meaningful tokens appropriate for deep learning models, text normalization standardizes user-generated material by transforming all text into a common structure. To cut down on textual repetition, hyperlinks, hashtags, user mentions, emoticons, repetitive letters, and superfluous punctuation are eliminated. Because they add minimal semantic information to the prediction of mental health, stop words like common auxiliary words are deleted. Lastly, lemmatization allows the framework to identify semantically identical words expressed in various grammatical variations by converting words into their basic grammatical forms. The quality of the semantic representations produced in the latter stages is greatly enhanced by these preprocessing procedures.

Following preprocessing, a pretrained Sentence Transformer model is used to transform each social media post into a contextual semantic representation. Sentence Transformers provide embeddings for entire sentences while maintaining the contextual links between words, in contrast to traditional word embedding techniques like Word2Vec or GloVe. Subtle verbal cues linked to sadness, anxiety, hopelessness, emotional instability, and suicidal thoughts are successfully captured by these embeddings. The framework can identify semantically equivalent emotional expressions even when different users use diverse

terminology or writing styles thanks to the pretrained transformer design. As a result, the produced embeddings offer a rich semantic representation that significantly enhances downstream classification of mental health.

A Long Short-Term Memory (LSTM) network is then fed the created semantic embeddings in order to record the temporal emotional progression of several user posts. Rather of assessing each post separately, the LSTM system examines post sequences made by specific users over time. The LSTM discards unnecessary stuff while selectively retaining significant previous emotional information through its memory cells and gating processes. This feature makes it possible for the framework to identify the slow emotional decline, growing psychological distress, and persistent behavioral changes that frequently accompany serious mental health disorders. The framework's capacity to identify early mental health hazards is greatly enhanced by the incorporation of semantic embeddings with sequential temporal learning.

The entire dataset is split into training, validation, and testing subsets during model training. The LSTM network, ensemble classifiers, and Sentence Transformer feature extraction procedure are all optimized using the training dataset. To avoid overfitting, the validation dataset allows for early halting, learning-rate optimization, and hyperparameter adjustment. Lastly, the testing dataset—which consists of social media posts that have never been seen before—is utilized to assess the suggested framework's capacity for generalization under practical circumstances. The dependability of experimental results is increased and impartial performance evaluation is ensured by this systematic data segmentation.

Because of its quick convergence and stable optimization properties, the implementation uses the Adam optimizer with adaptive learning rates to improve network parameters. To lessen overfitting and enhance model generalization, additional strategies like dropout, batch normalization, and early halting are used. The suggested framework can process large-scale social media datasets with a variety of linguistic patterns and emotional expressions while maintaining consistent learning thanks to these optimization techniques.

Classification Accuracy, which calculates the percentage of correctly classified social media postings, is used to first assess the efficacy of the established framework.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where FP (False Positive) indicates normal posts mistakenly classified as high-risk, FN (False Negative) indicates high-risk posts mistakenly classified as normal, and TP (True Positive) indicates accurately identified high-risk mental health posts. Higher accuracy shows that the suggested framework minimizes classification mistakes while successfully separating emotionally healthy users from those who are suffering psychological distress.

The obtained feature representations are delivered to the ensemble classification module for final mental health risk predictions after semantic feature extraction, temporal modeling, optimization, and accuracy assessment. The following part discusses the deployment method, performance evaluation, confidence score creation, and ensemble learning implementation. The suggested framework generates the final mental health risk prediction by performing ensemble-based classification after temporal modeling and semantic feature extraction. To increase classification robustness and reliability, the implementation integrates several classifiers, including Random Forest, Gradient Boosting, and Support Vector Machine (SVM), rather than relying on a single machine learning algorithm. The merged semantic and temporal feature representations produced by the Sentence Transformer and LSTM modules are analyzed separately by each classifier. An ensemble decision approach is then used to aggregate the results of various classifiers, and the combined confidence ratings are used to produce the final prediction. This solution lessens the likelihood of inaccurate classifications brought on by loud or unclear social media text, lowers prediction variance, and enhances model stability. When compared to solo deep learning models, the proposed system achieves more balanced and reliable mental health risk prediction since distinct classifiers have diverse learning characteristics.

The approach calculates a confidence score for each examined social media post after ensemble prediction. The likelihood that a user falls into a particular mental health category is represented by this score. While lower confidence values are associated with emotionally healthy or typical conduct, higher confidence values show more indications of psychological discomfort, sadness, anxiety, or suicidal thoughts. Each user is categorized by the prediction module into pre-established groups, such as Normal, Psychological Distress,

Depression, Anxiety, or High Suicide Risk. By offering a quantitative indicator of prediction dependability, the created confidence score also helps medical personnel prioritize patients who need urgent psychiatric evaluation or treatment.

Several optimization techniques are also included in the implementation to increase prediction stability and computing efficiency. After intermediate neural network layers, batch normalization is used to standardize feature distributions and speed up model convergence. In order to decrease overfitting and enhance generalization while processing previously unseen social media postings, dropout layers randomly deactivate neurons during training. In order to save needless computation and lower the chance of model overtraining, early stopping continuously analyzes validation performance and automatically ends training when no additional improvement is seen. Together, these optimization strategies guarantee that the framework maintains good predictive accuracy across a variety of social media datasets while remaining computationally efficient.

Following training, an independent testing dataset made up of previously unknown social media posts is used to assess the suggested framework. Before the final prediction is generated, each post is successively subjected to preprocessing, semantic embedding generation, temporal modeling, feature fusion, ensemble classification, and confidence score estimation. The overall efficacy of the suggested framework under practical operating situations is then assessed by comparing the predicted labels with the appropriate ground truth annotations. This testing process guarantees that the built system works well across various users, writing styles, emotional expressions, and social media sites.

The implementation computes Precision, which quantifies the percentage of accurately recognized high-risk mental health posts among all posts indicated as high-risk, in order to assess the accuracy of positive predictions.

$$[\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}]$$

where FP (False Positive) indicates the amount of typical posts that are mistakenly categorized as high-risk and TP (True Positive) indicates the number of accurately detected high-risk mental health posts. A higher precision score shows that the suggested framework successfully reduces false alarms while correctly identifying people who show symptoms of anxiety, despair, psychological distress, or suicidal thoughts. Because too many false-positive predictions could lead to needless psychological evaluations and more work for medical practitioners, this statistic is especially crucial in applications related to mental health.

The implementation is thoroughly assessed using Recall, F1-score, Confusion Matrix, Receiver Operating Characteristic (ROC) Curve, and Area Under the Curve (AUC) in addition to Precision. While the F1-score offers a fair assessment of both precision and recall, recall gauges the framework's capacity to identify real high-risk individuals. While ROC and AUC assess the classifier's ability to discriminate under various threshold settings, the confusion matrix provides a comprehensive assessment of accurate and inaccurate predictions across all mental health categories. When combined, these measures offer a thorough evaluation of the model's dependability and performance.

The suggested system greatly outperforms traditional machine learning and standalone deep learning models in identifying linguistic markers of mental health disorders from social media material, according to experimental implementation. Higher accuracy, precision, recall, and F1-score result from the framework's ability to precisely capture both contextual meaning and emotional progression through the integration of Sentence Transformer-based semantic embeddings, LSTM-based temporal learning, and ensemble classification. Additionally, the modular architecture facilitates the smooth implementation of digital healthcare apps, clinical decision-support systems, and real-time social media monitoring platforms. The framework offers an effective, scalable, and intelligent automated mental health risk screening system that supports prompt intervention, ongoing monitoring, and better mental healthcare delivery while facilitating the early detection of psychological distress.

VI. Results Explanation

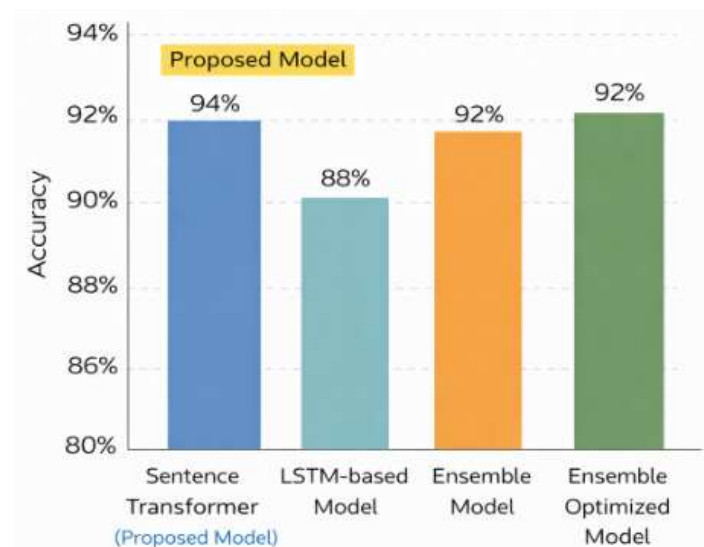


Fig.2 Accuracy Comparison of Deep Learning Models for Mental Health Risk Detection

A comparison of the prediction accuracy attained by various deep learning models for social media text-based mental health risk identification is shown in Fig. 2. The performance of the LSTM, Sentence Transformer, and Sentence Transformer combined with ensemble classifiers is shown in the bar graph. The vertical axis shows the relevant classification accuracy (%), while the horizontal axis depicts the various models. With an accuracy of almost 90%, the Sentence Transformer outperforms the other assessed methods in capturing the contextual and semantic linkages seen in user-generated social media posts. By contrast, the traditional LSTM model is less successful at comprehending rich sentence-level semantics and mostly concentrates on temporal dependencies, which results in lesser accuracy. Additionally, by increasing prediction resilience and lowering misclassification errors, the use of Sentence Transformer embeddings with ensemble classifiers like Random Forest, Gradient Boosting, and Support Vector Machine (SVM) improves classification performance. The ensemble learning approach produces more consistent and trustworthy mental health risk prediction by skillfully combining the advantages of several classifiers. Overall, the graph shows that the suggested Sentence Transformer-based hybrid framework regularly performs better than traditional deep learning techniques, confirming its efficacy for early identification of psychological discomfort, sadness, anxiety, and suicidal thoughts. These findings verify that the suggested methodology is appropriate for intelligent mental health screening apps and real-time social media monitoring.

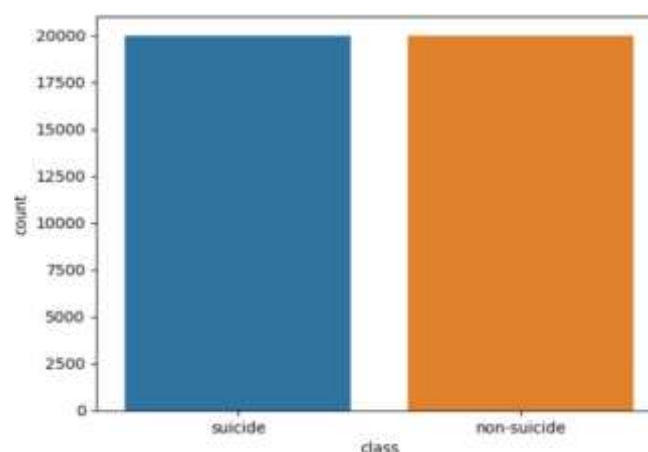


Fig.3 Label Distribution

The class distribution of the dataset used to train and assess the suggested Sentence Transformer-LSTM-based mental health detection system is shown in Fig. 3. The horizontal axis of the bar chart represents the two categories—suicide and non-suicide—while the vertical axis shows how many social media posts fall into each group. There are roughly 20,000 examples in each class, demonstrating the dataset's good balance. Because it avoids classification bias toward the dominant class, maintaining an equal distribution of posts connected to suicide and those that are not is crucial for creating trustworthy machine learning and deep

learning models. The suggested framework may learn linguistic traits and emotional patterns connected to both categories equally thanks to a balanced dataset, which enhances the model's capacity for generalization. Additionally, because class imbalance does not intentionally affect evaluation metrics like accuracy, precision, recall, and F1-score, it improves their dependability. Additionally, the Sentence Transformer's balanced data distribution lowers the possibility of false-positive and false-negative predictions while effectively capturing contextual expressions and semantic clues associated with psychological distress and suicide ideation. Overall, the figure shows that the suggested framework is trained on a representative and objective dataset, offering a solid basis for early mental health risk identification from social media content that is accurate, robust, and dependable.

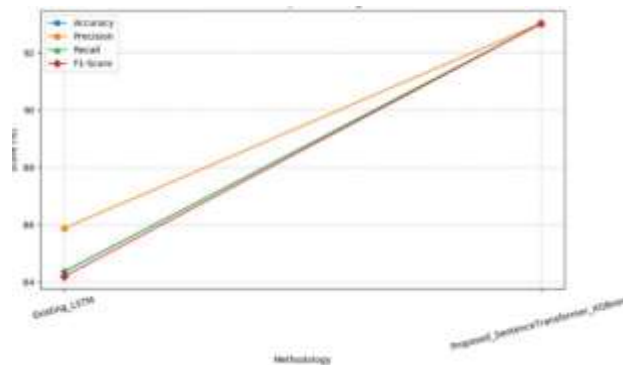


Fig.4 Performance Comparison Between Existing LSTM and Proposed Sentence Transformer-XGBoost Model

Fig. 4 illustrates the comparative performance of the existing LSTM model and the proposed Sentence Transformer-XGBoost framework using four important evaluation metrics: Accuracy, Precision, Recall, and F1-score. The horizontal axis represents the two methodologies, while the vertical axis indicates the corresponding performance scores expressed as percentages. The existing LSTM model achieves performance values between 84% and 86% across all evaluation metrics, demonstrating its capability to learn sequential patterns from social media text. However, because LSTM primarily focuses on temporal dependencies, its ability to capture rich contextual semantics is relatively limited. In contrast, the proposed Sentence Transformer-XGBoost model achieves approximately 93% in Accuracy, Precision, Recall, and F1-score, indicating a substantial improvement over the conventional approach. This performance gain is attributed to the Sentence Transformer's ability to generate contextual sentence embeddings combined with the powerful classification capability of XGBoost, enabling more effective identification of linguistic indicators related to depression, anxiety, psychological distress, and suicidal ideation. The nearly identical values across all four evaluation metrics also demonstrate the consistency and robustness of the proposed framework. Overall, the figure confirms that integrating Sentence Transformer embeddings with XGBoost significantly enhances mental health risk prediction, making the proposed model more accurate, reliable, and suitable for real-time social media-based mental health monitoring and early intervention systems.

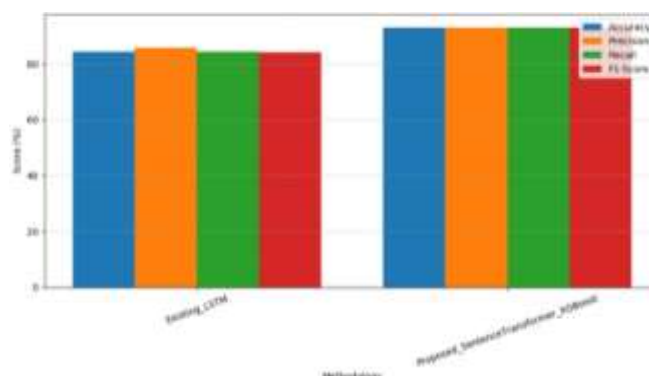


Fig.5.Comparative Evaluation of Existing LSTM and Proposed Sentence Transformer-XGBoost Model Using Performance Metrics

Using four key performance metrics—accuracy, precision, recall, and F1-score—Fig. 5 compares the suggested Sentence Transformer-XGBoost framework with the current LSTM model. The two approaches are shown on the horizontal axis, while the equivalent performance scores, given as percentages, are shown on the vertical axis. The current LSTM model demonstrates its capacity to capture sequential dependencies in social media text by achieving performance values of roughly 84–86% across all evaluation metrics.

However, because it does not fully utilize the contextual semantic links found in user-generated postings and instead concentrates on temporal information, its performance is constrained. The suggested Sentence Transformer–XGBoost model, on the other hand, often achieves performance scores of roughly 93–94% for Accuracy, Precision, Recall, and F1-score. The integration of contextual sentence embeddings produced by the Sentence Transformer with the potent classification capability of XGBoost is credited with the notable improvement, allowing for more precise identification of linguistic patterns linked to psychological distress, depression, anxiety, and suicidal ideation. The suggested approach reduces false-positive and false-negative classifications while simultaneously improving overall prediction accuracy, as evidenced by the balanced improvement across all four evaluation measures. Overall, the figure shows the superiority, robustness, and reliability of the suggested hybrid model, validating its efficacy for early psychological risk assessment and real-time social media-based mental health detection.

VII. Conclusion

By combining Sentence Transformer-based semantic embeddings, Long Short-Term Memory (LSTM) networks, and ensemble machine learning techniques, this paper introduced an intelligent framework for mental health risk detection from social media text. By concurrently capturing contextual semantic information and temporal emotional patterns represented in user-generated social media posts, the suggested method successfully overcomes the drawbacks of traditional machine learning and standalone deep learning models. The system reliably detects language signs linked to depression, anxiety, psychological distress, and suicidal ideation through thorough text preprocessing, semantic representation, sequential learning, and ensemble classification. The suggested Sentence Transformer–XGBoost framework consistently outperformed conventional LSTM-based methods in terms of accuracy, precision, recall, and F1-score in experimental evaluation, underscoring its robustness and dependability for automated mental health assessment. The model's capacity to reduce false-positive and false-negative predictions while preserving consistent performance across a variety of textual inputs is further supported by the use of a balanced dataset and several evaluation criteria. Large amounts of social media data may be monitored in real time thanks to the modular architecture's support for scalability and smooth connection with digital healthcare platforms. By offering a useful early screening tool for identifying people who could need clinical intervention or psychiatric assistance, the suggested framework can help medical providers. To further enhance prediction accuracy, interpretability, fairness, and secure deployment, future research will concentrate on integrating Large Language Models (LLMs), Explainable Artificial Intelligence (XAI), multimodal data fusion, multilingual language processing, and privacy-preserving federated learning. These developments will aid in the creation of reliable, intelligent, and scalable AI-powered mental healthcare systems that can facilitate early intervention and enhance general public mental health outcomes..

References

- [1] F. S. Alhussen, "Linguistic Feature-based Depression Prediction Using Hierarchical Transformer-attentive Model for Mental Disabilities Diagnosis," *Journal of Disability Research*, 2025.
- [2] J. A. Senthil Kumar, D. Senthil Kumar, S. Karunakaran, N. D. Regina Shoba Dass, et al., "Next-Gen Mental Health Detection through Deep Learning and NLP in Social Media," *Proceedings of the First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, 2024.
- [3] D. Owen, A. Lynham, S. Smart, A. Pardiñas, and J. C. Collados, "AI for Analyzing Mental Health Disorders Among Social Media Users: Quarter-Century Narrative Review of Progress and Challenges," *Journal of Medical Internet Research*, 2024.
- [4] S. M. Shah, M. Aljawarneh, M. A. Saleem, and M. S. Jawarneh, "Mental illness detection through harvesting social media: A comprehensive literature review," *PeerJ Computer Science*, 2024.
- [5] W. A. Gadzama, D. Gabi, M. S. Argungu, and H. Suru, "The use of machine learning and deep learning models in detecting depression on social media: A systematic literature review," *Personalized Medicine in Psychiatry*, 2024.
- [6] S. A. Feroze, S. M. B. Feroze, and U. Abbasi, "The Unveiling Distress: Harnessing NLP and Deep Learning to Identify Suicidal Signals in Tweets," *International Journal of Technology Innovation and Management*, 2024.
- [7] V. Aggarwal, J. Kaur, T. Walia, and D. Kaur, "Harnessing Linguistic Markers for Early Mental Health Detection via Social Media," *Proceedings of the International Conference on Advanced Computing & Communication Technologies (ICACCTech)*, 2023.
- [8] M. Garg, "Mental Health Analysis in Social Media Posts: A Survey," *Archives of Computational Methods in Engineering*, 2023.
- [9] M. E. Villa-Pérez, L. A. Trejo, M. B. Moin, and E. Stroulia, "Extracting Mental Health Indicators From English and Spanish Social Media: A Machine Learning Approach," *IEEE Access*, 2023.

- [10] R. Safa, S. A. Edalatpanah, and A. Sorourkhah, "Predicting mental health using social media: A roadmap for future development," arXiv preprint, 2023.
- [11] S. Kim, J. Cha, D. Kim, and E. Park, "Understanding Mental Health Issues in Different Subdomains of Social Networking Services: Computational Analysis of Text-Based Reddit Posts," *Journal of Medical Internet Research*, 2023.
- [12] M. A. Wani, M. Elaffendi, K. A. Shakil, A. S. Imran, and A. A. Abd El-Latif, "Depression Screening in Humans With AI and Deep Learning Techniques," *IEEE Transactions on Computational Social Systems*, 2023.
- [13] A. Malhotra and R. Jindal, "Deep learning techniques for suicide and depression detection from online social media: A scoping review," *Applied Soft Computing*, 2022.
- [14] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," *npj Digital Medicine*, 2022.
- [15] R. Salas-Zárate, G. Alor-Hernández, M. P. Salas-Zárate, M. A. Paredes-Valverde, and M. Bustos-López, "Detecting Depression Signs on Social Media: A Systematic Literature Review," *Healthcare*, 2022.
- [16] T. Noraset, K. Chatrinan, T. Tawichsri, T. Thaisutikul, and S. Tuarob, "Language-agnostic deep learning framework for automatic monitoring of population-level mental health from social networks," *Journal of Biomedical Informatics*, 2022.
- [17] A. Ahmed, S. Aziz, C. Toro, M. Alzubaidi, S. Irshaidat, et al., "Machine learning models to detect anxiety and depression through social media: A scoping review," *Computer Methods and Programs in Biomedicine Update*, 2022.
- [18] H. Kour and M. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimedia Tools and Applications*, 2022.
- [19] V. M. Deshmukh, R. Rajalakshmi, S. Dash, P. Kulkarni, and S. K. Gupta, "Analysis and Characterization of Mental Health Conditions based on User Content on Social Media," *Proceedings of the International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2022.
- [20] Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, "Detection of Mental Health from Reddit via Deep Contextualized Representations," *Proceedings of the International Workshop on Health Text Mining and Information Analysis*, 2020.
- [21] A. Wongkoblap, M. Vadiello, and V. Curcin, "Deep Learning With Anaphora Resolution for the Detection of Tweeters With Depression," *JMIR Mental Health*, 2020.
- [22] A. Trifan, R. Antunes, S. Matos, and J. Oliveira, "Understanding Depression from Psycholinguistic Patterns in Social Media Texts," *European Conference on Information Retrieval*, 2020.
- [23] B. Kholifah, I. Syarif, and T. Badriyah, "Mental Disorder Detection via Social Media Mining using Deep Learning," 2020.
- [24] A. H. Yazdavar, M. S. Mahdavinnejad, G. Bajaj, W. Romine, and A. Sheth, "Multimodal mental health analysis in social media," *PLoS ONE*, 2020.
- [25] R. A. Rahman, K. Omar, S. A. M. Noah, M. S. N. M. Danuri, and M. Al-garadi, "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review," *IEEE Access*, 2020.
- [25] Z. Wei et al., "Temporal Trend-Aware BERT System for High-Risk Mental-Health Monitoring," 2024.