International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 6, 2146-2170, 2025
https://doi.org/10.33889/IJMEMS.2025.10.6.100

Ram Arti Publishers

DDNet: A Novel Hybrid Deep Learning Model for Detection and Classification of Depression in Social Media Conversations

Md Zainuddin Naveed
Department of Computer Science and Engineering,
GITAM School of Technology Hyderabad, GITAM (Deemed to be University), Hyderabad Campus Rudraram,
Patancheru Mandal, Sangareddy, Hyderabad, 502329, Telangana, India.
Corresponding author: mnaveed2@gitam.in

Shivampeta Aparna
Department of Computer Science and Engineering,
GITAM School of Technology Hyderabad, GITAM (Deemed to be University), Hyderabad Campus Rudraram,
Patancheru Mandal, Sangareddy, Hyderabad, 502329, Telangana, India.
E-mail: ashivamp@gitam.edu

(Received on December 9, 2024; Revised on February 22, 2025 & May 6, 2025 & June 13, 2025; Accepted on July 5, 2025)

Abstract
In the modern world, people worldwide face different forms of depression due to factors such as workplace stress, economic pressures, and other causes. The rise of Artificial Intelligence (AI) has enabled data analysis and solving of real-world problems. People frequently use social media platforms to communicate and express their feelings. Hence, social media data is helpful for research purposes, particularly for automatic depression detection. Numerous scholarly works have explored using learning-based approaches to identify sadness from social media interactions. However, individual existing deep learning models have limitations, such as the inability to capture contextual and sequential dependencies in text fully. We addressed this by proposing a deep learning-based, non-invasive approach to identify depression in social media conversations. Our proposed approach involves a novel hybrid deep learning model, Depression Detection Network (DDNet), which combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models. The model was trained and tested on a manually annotated dataset of 8500 depression-related tweets (6,800 for training and 1,700 for testing) collected via the Twitter Application Programming Interface (API). The DDNet model achieved a high accuracy of 96.21%, outperforming baseline models such as standalone LSTM (92.31%) and Recurrent Neural Network (RNN) (91.43%). Furthermore, we developed Hybrid Deep Learning-based Depression Detection (HDL-DD), an algorithm that processes social media text and predicts potential depressive tendencies. The experimental results indicate that DDNet significantly improves depression classification, achieving 95% precision, 96% recall, and 95% F1-score, demonstrating its effectiveness over existing methods. By recognizing depression with a 96.21% accuracy rate, our deep learning model outperformed previous state-of- the-art approaches, making it a promising tool for automated depression monitoring applications. This approach could be integrated into real-world social media-based mental health monitoring applications, supporting early intervention efforts and contributing to AI-driven healthcare solutions.

Keywords– Depression detection, Artificial intelligence, Deep learning, Hybrid deep learning, Online social media.

1. Introduction

Millions of individuals worldwide suffer from depression, a prevalent mental disorder. Social pressures, personal problems, stress at work, and financial difficulties are some of the things that might cause depression. A person's life may be significantly impacted by depression if treatment is not received, resulting in ongoing sadness or unhappiness and chronic pain. It may also affect one's overall functioning, decision-making, and concentration. Furthermore, the psychological ramifications of depression may result in physiological complications. To diagnose depression, researchers have been developing methods based on artificial intelligence. Multi-modality techniques that consider many data sources can be employed to assess the probability and severity of depression (Salvi et al., 2024). Early identification and the process of implementing the required strategies to assist patients in overcoming depression can be aided by Artificial

Intelligence (AI)-based approaches (Cascarano et al., 2023). Automated depression identification in clinical settings has improved with Deep Learning (DL) and Machine Learning (ML) approaches. Analyzing data from several sources improves the detection procedure's accuracy (Thati et al., 2023). When used on neuroimaging samples, deep learning models and generative adversarial networks should produce helpful new information for a technology-driven method of diagnosing schizophrenia. This method gives fresh perspectives on depression and its treatment (Wang et al., 2023; Wang et al., 2024).

Depression, a significant mental health disorder suffered by millions of people around the globe, is complex to detect, as identifying it is inherently subjective, and there are no readily accessible diagnostic tools. Social networks have become an essential source of information for detecting trends in mental health based on how users interact with these platforms. Still, current deep learning models suffer from contextual awareness and sequential dependencies when discussing the text. In an effort towards closing this gap, we introduce a Differential Detection Network (DDNet), a novel hybrid deep learning model that combines the complementary strength of Convolutional Neural Networks (CNN) that perform feature extracting and Long Short-Term Memory (LSTM) networks, which are best performing sequential data learning framework. The main goals of this research are (1) designing a hybrid deep learning framework, which is an effective method to detect and classify depressive language in social media conversations, (2) experimentally verifying DDNet against best-in-class depression detection models, and (3) assessing the potential of exploration and implementation on automated mental health monitoring systems. Through these goals, our initial step will lead us toward our ultimate aim of implementing non-invasive AI-driven depression detection programs that could enhance early intervention strategies.

Numerous efforts have been made to tackle this issue but they are challenged by the high heterogeneity of text due to slang and sarcasm and temporal aspects. In addition, most of the models are not explainable or not taught with domain specific data. To overcome these and similar gaps, we propose in this work a personalized DDNet architecture, learned in a controlled setting retrieving a Twitter-annotated dataset, for the purpose of achieving better interpretability and robustness for early depression detection.

Our work provides a DL-based, non-invasive method for identifying sorrow in social media talks. DDNet is a novel hybrid DL model that blends LSTM and CNNs. Based on our empirical investigation, DDNet demonstrated enhanced depression detection capabilities. Furthermore, using the suggested deep learning model, we developed an HDL-DD algorithm that can examine social media posts for signs of depression 96.21% of the time; our deep learning model correctly diagnosed depression, which is good. Our suggested approach might be included in applications that automatically scan social media chats for signs of depressive illness. The rest of the document is organized as follows: Section 2 looks at the study of the various methods currently used to identify Sadness and social media statistics automatically. The suggested approach, intended to use Section 3, presents the use of social media interactions for the automated detection of sadness. In Section 4, we present the results of our empirical analysis. Section 5 discusses the work done in this paper along with the study's limitations, while Section 6 focuses on social and practical implications. Section 7 wraps up the study conducted for this publication and provides potential directions for further research.

2. Related Work

Many existing approaches use deep learning to find people's mental health based on conversations. Salvi et al. (2024) addressed the challenges of integrating multimodality approaches to diagnosis and prognosis in healthcare while emphasizing their benefits, which include more effective therapy and customized medicine. While modalities have different qualities, AI helps with data fusion; ethical and technological challenges must be addressed in future research. Cascarano et al. (2023) emphasized the benefits of using

ML for longitudinal biological data to facilitate early diagnosis and therapy planning. The hurdles are in the data complexity and model building. Developing algorithms for clinical applications and addressing machine learning interpretability issues will be the main focus of future research projects. Thati et al. (2023) diagnosed depression by combining task-based and mobile crowd-sensing methods with machine learning and multimodal data. The 86% accuracy rate of SVM demonstrates the benefits of feature fusion; nonetheless, interpretability and real-world application remain challenges. Future initiatives will aim to improve feature selection and responsibly manage ethical data. Wang et al. (2023), while recognizing problems with repeatability and interpretability, the study highlighted the usefulness of GANs in brain research and neuroimaging for disease detection. Future work will concentrate on improving clinical integration and algorithm transparency. Wang et al. (2024) used MSFNet to classify MI-EEG with high accuracy and F1 scores. Future studies might examine larger datasets and address realistic limitations in real-world applications.

Nash et al. (2023) explored ML applications for fMRI and EEG, among additional techniques, in the detection of depression and ADHD. Subsequent research will focus on establishing collaborative multimodal databases while considering privacy and dataset size constraints. Huang et al. (2024a) investigated the use of deep learning to integrate explainable AI (XAI) with interpretable AI (IAI) in healthcare natural language processing—subsequent research endeavors to refine worldwide modeling and formulate optimal methodologies for resilient XIAI implementation. Ahmed et al. (2022) presented an emotional lexicon and a structural hypergraph for NLP word representation. Upcoming tasks involve enhancing the choice of embeddings and evaluating the model's scalability to improve dependability. Sujith et al. (2022) examined the developments in intelligent health monitoring (SHM) made possible by blockchain, artificial intelligence, and 5G. Extending SHM in sports and individualized treatment are the focus of future research. Huang et al. (2024b) provided fair clinical decision assistance; the study examines AI/ML fairness in healthcare and identifies research needs.

Shah et al. (2020) used a hybrid approach to improve the identification of sadness from social media posts. Good classification using Word2VecEmbed + Meta characteristics is one of its merits; nevertheless, long detection times need to be improved. Zogan et al. (2022) aimed to improve MDHAN for depression identification by concurrently assessing subjects and attitudes in users' tweets. William & Suhartono (2021) focused on enhancing (Bidirectional Encoder Representations from Transformers) BERT- based models for depression identification by improving summarization methods and hyperparameter adjustments. Vandana et al. (2023) used a CNN-biLSTM hybrid model to analyze Twitter data and predict depression with reasonable accuracy. To achieve better performance, future studies will examine pre-trained models like BERT and optimize neural network setups. Ding et al. (2020) used deep neural networks and DISVM to identify depression in college students based on Sina Weibo data. Improved objectivity in feature selection and efficient data selection are areas of focus for future research.

Uddin et al. (2022) created a successful LSTM-based RNN to identify signs of sadness in text from a Norwegian youth information channel. Future efforts will incorporate explainable AI for decision- making and extend it to more languages. Yang et al. (2022) proposed KC-Net as a technique that uses mental state data and contrastive learning to detect stress and depression on social media. Among the benefits are creative results. The subsequent research aims to improve the detection of stress factors and information integration. Uban et al. (2021) discussed DL models while researching the identification of mental disorders using social media language analysis. Enhancing emotion-temporal analyses is the focus of future research. In his study of social media for mental health prediction, Garg et al. (2023) focused on AI models and dataset availability. Longitudinal suicide risk analysis will be a part of future efforts. Safa et al. (2022) aimed to use social media data to predict sadness by utilizing many features and classifiers. To enhance the accuracy of mental health diagnoses, future research will concentrate on improving feature selection and incorporating clinical data.

Rahman et al. (2020) provided a critical overview of the approaches used in machine learning and data analysis to detect mental health issues in online social networks. Future research will focus on enhancing accuracy and tackling issues like data privacy. With improved datasets and algorithms, Alghamdi et al. (2020) aim to increase accuracy and model complexity by utilizing ML and NLP to investigate Arabic social media for depression prediction. Murshed et al. (2022) developed the DEA- RNN model, which outperforms the Bi-LSTM, RNN, SVM, MNB, and RF models for Twitter cyberbullying detection. Multimedia analysis integration and platform expansion are tasks for the future. Sanchez et al. (2020) examined ML strategies for identifying suicide on social media, with a focus on validation measures, data sources, and methodology. Future research will focus on improved machine learning models and multilingual studies. Cao et al. (2020) found that suicide intentions on social media can be detected by using a personal knowledge graph and a two-layered attention mechanism. An accuracy above 93% was attained, but continuous knowledge graph upkeep and data dependability remain issues.

Rao et al. (2020) suggested hierarchical models (MGL-CNN, SGL-CNN) perform better in detecting sadness on social media, handling vast amounts of data presents difficulties. More general uses of sentiment analysis are planned for future study. Li et al. (2020a) proposed that the HEMOS system incorporates vocabulary, comedy, and emojis to improve sentiment categorization in Chinese. Plans for the future include adding graphics, dealing with spam posts, and increasing the amount of labeled data. Gedam and Paul (2021) investigated wearable sensor-based stress detection using machine learning methods. In the future, improve accuracy, fix issues quickly, and increase reliability. Wang et al. (2020) used Weibo to evaluate COVID-19 sentiment using BERT and TF-IDF. Plans include expanding the quantity, improving the precision of the models, and using communication platforms. Xu et al. (2022) looked into sentiment analysis in social media, highlighting the obstacles and potential avenues for further multilingual study and incorporating a larger corpus of literature.

Amin et al. (2022) Subsequent studies will expand the symptoms' spectrum and enhance model performance by applying advanced word embedding methodologies. To reach greater relevance, Al-Omari et al. (2020), future studies will examine complex deep learning systems and broaden the emotional spectrum. Khan et al. (2022) future development is to make HCovBi-Caps more scalable across several social media networks and resistant to hostile attacks. Alsubhi et al. (2023) improved accuracy and diversified the AraBig5 dataset; future research will use advanced deep-learning methods. Grover et al. (2022) investigated social media scrutiny, integration of additional professional perspectives, and empirical data verification of assertions.
Li et al. (2020b) enhanced social media data preprocessing, developing crisis effect prediction models, and perfecting algorithms that foresee people's thoughts and actions during power outages will all be necessary. Elsafoury et al. (2021) examined BERT pretraining on slang-based writing to enhance cyberbullying detection effectiveness. Amjad et al. (2021) subsequent actions involve expanding the list of hazardous Urdu work and utilizing transformer models to enhance identification. Kumar et al. (2021) future studies will leverage advanced word embeddings like BERT to improve performance on clinical data classification challenges. Yue et al. (2020) concentrated on balancing participant demographics, growing databases for broader validation, and looking into other population-specific elements to increase the accuracy of depression screening using Internet traffic.

From the review above, it is observed that deep learning algorithms have enhanced the level of accuracy in the area of sentiment and emotion classification, but models are not neatly optimized to capture the same sentiment and response in depression detection in social media. This is the motivation of the proposed

DDNet model, which aims at learning context aware feature for better detection performance. While deep learning models are increasingly being employed for depression detection, they all have certain shortcomings. Traditional machine learning methods like SVM, Random Forest, etc., which depend extensively on human-designed features, fail to capture the rich characteristics of depressive statements. While deep learning models, in particular, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have achieved better performance in text-based sentiment analysis, researchers have also found that CNNs alone have significant limitations in capturing long-range dependencies in textual data. RNNs, including LSTM, face the problems of vanishing gradients and high computational costs.

Additionally, previously employed integrated models fail to achieve sufficient accuracy for accurately detecting nuanced Depressive Expressions in social media conversation due to limited integration mechanisms between spatial feature extraction (CNNs) and sequential context learning (LSTMs) methods. We introduce DDNet, a hybrid deep learning model that successfully fuses CNN and LSTM layers to address these issues. The CNN part of the model augments local feature extraction to discover critical patterns correlated to depression. In contrast, the LSTM aspect of the model learns the sequential dependencies of the text, enabling it to accommodate the contextual meaning contained in conversations. This combined method helps the model better distinguish depressive content from general sentiment relative to purely sentiment-based models, thus making it a stronger candidate for practical applications in depression detection.

3. Materials and Methods

This section describes the resources and techniques utilized in the research described in this paper. It outlines the suggested deep learning framework's workings and specifics about the dataset used, the proposed hybrid DL model, and the methodology for its evolution.

3.1 Problem Definition

Creating a DL framework and algorithm to recognize sad behavior in people instantly is a significant problem. This issue is critical in the current environment because of several reasons: pressures from the workplace, society, lifestyle, and finances. Since social media platforms have become famous, individuals from all walks of life freely express their thoughts and emotions. These platforms offer abundant data for studies that seek to comprehend people's mental health. In this study, the problem is modeled as a binary classification task, where each social media post is labeled as either depressive or non-depressive based on manual annotation.

3.2 Proposed Framework

Figure 1 illustrates the suggested architecture for detecting depression in social media interactions. It searches for signs of depression by filtering raw Twitter data using a mixed DL method. The system's original Twitter dataset sometimes includes duplicate or unneeded data, which might lead to increased noise and poorer model performance. The first step in resolving the problem is eliminating duplicate tweets, which improves the dataset by making it more representative and redundant. This step is necessary to ensure that the subsequent study focuses on specific data points, improving the results' relevance and quality. Our work offers a non-invasive, DL-based approach to detect sadness in social media conversations. A new hybrid deep learning model called DDNet combines CNNs and LSTM. Our empirical study showed that DDNet had improved depression detection skills. Additionally, we developed an algorithm called HDL-DD that can recognize melancholy in social media postings using the proposed DL model. Our DL model achieved a high 96.21% accuracy rate in identifying depression. Applications that automatically search social media conversations for indications of depression may incorporate our recommended method.

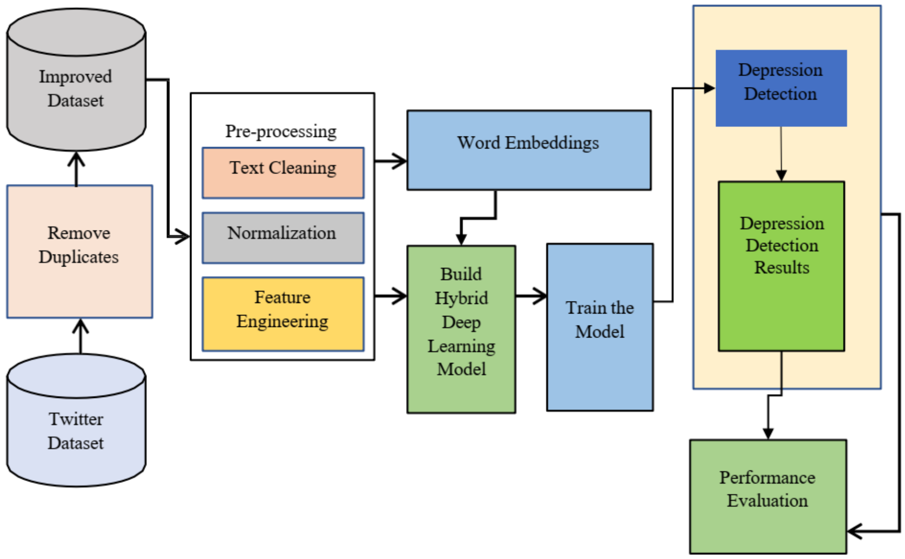

Figure 1. Proposed hybrid deep learning framework for depression detection from social media conversations.

Following preprocessing, word embeddings transform the spoken input into illustrations of dense vectors. Semantic connections among words are captured by already-trained models, which help the DL model comprehend the context and emotion of each tweet. Word embeddings are essential because they can store word meanings and relationships in a numerical format that neural networks can understand. It helps identify subtle speech patterns and emotional indicators indicating sad language. CNN and LSTM networks are combined in a hybrid deep learning model that forms the system's foundation. This hybrid method is ideal for text data analysis since the CNN component may identify local characteristics, including particular word or phrase patterns that could suggest a negative mood. In addition to allowing the model to comprehend the context and sequential flow of words, the LSTM element records temporal relationships. This is necessary to understand advanced emotive language structures. Pre-processed and embedded text data trains a combined DL model to recognize patterns that suggest a gloomy attitude. The algorithm can forecast the probability of depressive indications in Twitter correspondence after it has been taught. A tweet's likelihood of containing depressing material can be ascertained using the model's output, which represents the findings of the depression detection study. To ascertain whether the tweet signals depression, this detection phase entails introducing new data to the trained model and assessing the outcomes. Lastly, the system's performance evaluation evaluates the model's accuracy and efficacy.

Indeed, the novelty of the proposed DDNet framework is the capability of jointly leveraging the spatial feature extraction capability of CNNs and the temporal long sequence learning ability of LSTM networks. Our hybrid approach not only combines the power of CNNs and RNNs to understand the patterns in depressive language better but also allows us to capture both short-range local word dependencies and the long-range contextual relationships that are necessary for accurate classification. They combine the word embeddings with CNN to discover critical patterns indicative of depression at the phrase level. At the same time, the X layer allows the temporal dependencies for classification tasks to be stated, thus improving the accuracy of classifying the sentiment of the emotional language in social media.

The proposed model is adapted to process large-scale social media datasets to facilitate scalability by utilizing parallel processing strengths offered in deep learning functional libraries such as TensorFlow and PyTorch. This architecture enables incremental learning, allowing the model to adapt to changes in linguistic styles and vocabulary associated with depression over time, particularly in social media contexts. Moreover, the model can be adapted to work with multimodal inputs, integrating text with images or speech data to provide a more comprehensive understanding of a user's mental state. The development of AI-driven models for depression detection poses ethical challenges in terms of user privacy, data security, and potential biases in prediction outcomes. To address these concerns, our framework implements anonymization and ensures user data is processed as per ethical AI principles. These novel techniques can also be combined with existing explainability techniques like attention mechanisms and feature importance to improve model interpretability and overall reliability for real- world applications. And a visual of a hybrid deep learning pipeline, like a better block diagram of DDNet, should be added to showcase the architecture workflow, from data preprocessing, feature extraction, model training, and depression classification. Such advancements will augment our method from even more technical depth to practical applicability.

3.3 Proposed Hybrid Deep Learning Model

Using emotion analysis, we suggested a hybrid DL model to assess and identify symptoms of depression in tweets and other social media posts. In sequential text input, the CNN-LSTM network combination model is ideal for sentiment-based depression diagnosis because it can identify persistent connections and local feature patterns. The model uses a trained encoding layer to convert each word into a fixed-dimensional vector representation in the input text. Using a word-to-vector map that has already been built on extensive text datasets, like GloVe or Word2Vec, these embeddings are generated, denoted as x(t) for the t^{th} word in the input sequence. To improve the model's comprehension of the finer points of language—a critical component of sentiment analysis—the embedding layer encodes the semantic meaning of words into continuous-valued vectors. Where the embedding layer's dimensionality is denoted by d, it creates a matrix of word vectors with dimensions L×dL×d given a maximum sequence length L.

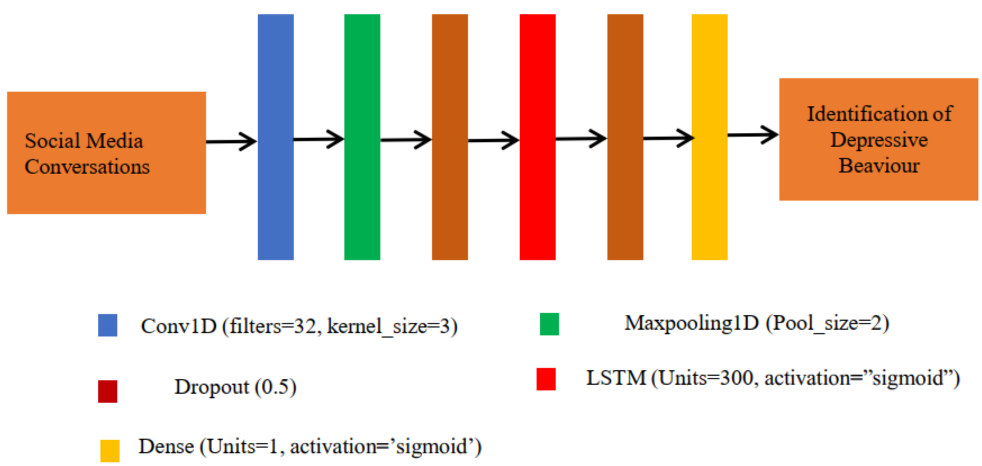

Figure 2. Proposed hybrid deep learning architecture, known as DDNet, made up of CNN and LSTM.

The one-dimensional convolutional layer (Conv1D), which has a kernel size of three and 32 filters, is applied after the embedding layer. This layer captures localized information, including particular word or phrase patterns that may suggest depressed language or bad attitude, by performing convolution operations on each subsequence of the word embeddings. In a sliding window method, the convolution process applies each filter W_f throughout the input sequence. The filter first calculates a weighted sum of word embeddings within the window for each location. Then, to add non-linearity, an activation function for ReLU is used.

In terms of mathematics, the convolution operation at the point it is represented as follows:

$$h_t = ReLU(W_f . x_{t:t+k-1} + b_f) \quad (1)$$

where, k is the kernel size, W_f is the filter, b_f is the bias term, and $x_{t:t+k-1}$ is the sequence of embeddings inside the window. The final product is a feature map that highlights significant n-gram features. In the CNN layer, the core operations can be expressed as Equation (1), where the filter is moved over input word vectors to filter them. This ensures non-linearity, enhancing learning while passing the weighted sum of inputs through the ReLU activation function, which allows extracting features from local text patterns related to depression detection.

Only the highest value in each window is retained when the feature map undergoes down sampling when the MaxPooling1D layer is applied with a pool size of two. This pooling technique not only increases computation performance but also lowers the size of the feature map to guarantee that the model concentrates on the most critical features inside each zone. Afterward, pooling the feature map reduces complexity while preserving essential patterns more straightforwardly. With a dropout rate of 0.5, a dropout layer comes after the max-pooling layer to avoid excessive fitting. In every training cycle, 50% of the neurons are arbitrarily set to zero. Dropout is a normalization that is better for adjusting the model to new, unknown input and preventing it from becoming unduly dependent on any one neuron.

To this end, a novel learning architecture, DDNet, combines the advantages of CNNs and LSTMs to facilitate depression detection in social media dialogue. To address this, we usually utilize the embedding layer, consisting of dense vector representations that contain information on the semantic relationship between words and allow for better contextualization. The Conv1D layer acts as a feature detector program local patterns in the input features, enabling the model to identify key depressive signals, while the max pooling layer works by making dimensional reduction and filtering noise out, which promotes computation efficiency. This question includes the LSTM layer because it helps to keep short- and long-term dependencies required to learn sequential patterns within the text, which is useful when distinguishing between just a negative sentiment versus actual depressive symptoms.

Although more recent methods, such as GRU, Transformer and BERT have achieved great success in various NLP tasks, we selected LSTM, which offers a good trade-off between accuracy and efficiency for sequential text classification, as it is lightweight and fast. However, fewer weighted gates may make GRUs less effective in capturing highly non-linear long-range dependencies, as found in depressive language. Transformers and BERT models require extensive pretraining and substantial computing resources, which hinders their deployment in real-time and resource-limited scenarios. In addition, BERT is developed explicitly for contextual embedding and is still less interpretable when directly applied to sequence modelling than LSTM's gating mechanism. Therefore, LSTM provides an interpretable and resource-friendly approach to learn significant contextual language patterns in social media conversations that might be associated with depression.

Dropout layers prevent overfitting, and some neurons are randomly set to zero during training to improve generalization. The last layer (Dense) connects features learned from the previous layers (features extracted in the classification process), aggregates them, and maps them to a binary output. Lastly, the sigmoid activation gives scores between 0 and 1, meaning that a more significant value will mean that the text is depressive. Based on a combination of CNN and LSTM, combining time invariance and temporal order, resulting in a feature extraction hybrid model resulting in 96.21% accuracy, which is higher than the use of standalone LSTM and CNN. This combination of structured layers facilitates efficient text representation, improving model interpretability and depression classification, which makes it an excellent choice for real-world applications in mental health monitoring and AI-driven intervention strategies. We select DDNet, a hybrid architecture with convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to extract social media text data's most powerful spatial and sequential features. Many existing models fail to strike a balance between feature extraction and contextual awareness, which restricts their ability to perceive depression accurately.

Justification of CNN in DDNet:

CNN has been incorporated into DDNet due to its capability to extract high-level features from text sequences. Leveraging convolutional filters, CNN captures local dependencies and phrase-level patterns, allowing it to learn critical depressive signals like negative sentiment, recurrence of emotional statements, and intensity of language. CNN's feature maps help apply noise filters, paying more attention to essential text structures, giving this model robustness to sentence length and word order.

Why LSTM in DDNet?

Although CNN captures useful spatial features, it cannot preserve sequential dependencies within long text sequences. Thus, to overcome this limitation, LSTM is added to the model, which keeps the memory of the previous words in a sequence using its gating mechanisms. This type of training allows the model to learn contextual changes, for example, progressive emotional decline or symptoms of depression that appear gradually over several sentences.

Why a Hybrid CNN-LSTM Model?

Stand-alone CNN does not consider word order beyond its receptive field, while a stand-alone LSTM usually performs poorly in selecting features. DDNet takes advantage of CNN's strength in extracting high-impact features and LSTM's advantage in modeling long-term dependencies, which is suitable for handling complex, sentiment-rich text in social media conversations. Furthermore, the CNN can also be a dimensionality reduction layer before LSTM, reducing computation overhead and preventing overfitting. This hybrid method improves the model's generalization, allowing DDNet to better distinguish between depressive and non-depressive conversations. Combining spatial feature extraction with sequential learning empowers DDNet to be a scalable and interpretable model for depression detection.

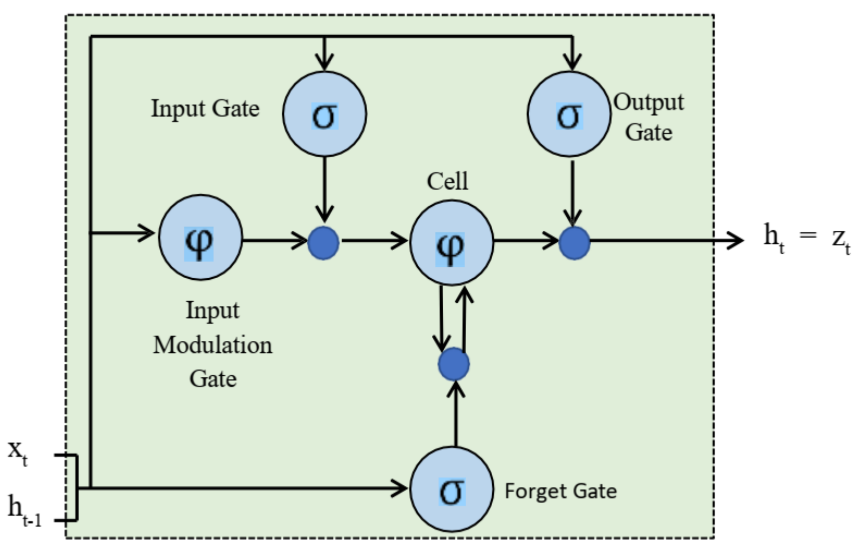

Figure 3. Architectural overview of LSTM unit.

Figure 3 Architectural overview of LSTM unit Time relationships and contextual information in the sequence are captured by adding a 300-unit LSTM layer after the convolutional and pooling layers. The identification of problematic phrases is made possible by the LSTM design's ability to efficiently describe the linguistic context and store information for extended periods. The internal structure of an LSTM cell, which helps the model follow the long-distance time steps within tweets of users, is depicted in **Figure 3**. This is especially important in the task of depression detection where emotional context and behavior cues might come from multiple tweets. The forget and input gates assist in filtering out unnecessary signals while keeping the depressive ones, thereby leading to superior performance in sequence modeling and classification efficiency, as demonstrated from the experimental results. The LSTM cell employs a combination of gating mechanisms to effectively regulate the flow of information and adapt the model to new inputs, as defined by the following equations:

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \tag{2}$$

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \tag{3}$$

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \tag{4}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_c . [h_{t-1}, x_t] + b_c) \tag{5}$$

$$h_t = o_t * \tanh() \tag{6}$$

where, i_t, f_t, and o_t represent the input, forget, and output gates, respectively; c_t the cell state, the hidden state by h_t and the hyperbolic tangent activation function by tanh. Both immediate and long-term dependencies are captured by the LSTM layer to comprehend the mood of entire patterns, even sentences that may imply depressing feelings, even if they occur later in the text. The LSTM is followed by a second dropout layer with a 0.5 dropout rate to further reduce the likelihood of overfitting. Equations (2)-(6) describe the central operations of an LSTM cell that allow you to keep long-term dependencies of text. An input gate decides the amount of new information to persist; a forget gate decides which data to throw away, and an output gate has a hidden state to produce. They allow effective sequence learning and help retain better context, crucial for depression detection.

The final method employs a single-unit dense layer with a sigmoid activation function. Based on this variable, the prototype will generate a probability value ranging from 0 to 1, reflecting the likelihood that the text contains sadness-related content. This layer provides the following definition for the sigmoid function:

$$y = \frac{1}{1+e^{-z}} .$$

When the input attributes from the layer before are linearly combined to create z. By applying the sigmoid function, the input logits are transformed into a probability ranging from 0 to 1, which is appropriate for binary classification problems like depression vs. non-depression in tweets. This function is differentiable and supports smooth gradient flow during training.

3.4 Algorithm Design

By employing state-of-the-art DL methods, the HDL-DD algorithm aims to accurately identify depressive conduct in social media interactions, specifically using Twitter data. Social media platforms contain vast amounts of unstructured data where users frequently express emotions, making them valuable for mental health analysis. However, detecting depression-related language in this data requires models that can interpret localized patterns (e.g., specific phrases) and sequential context (e.g., sentence flow), as depressive expressions are often nuanced and complex. The HDL-DD algorithm combines CNN and LSTM models in a hybrid architecture to capture these patterns effectively. By applying preprocessing, feature engineering, and embedding techniques, the algorithm prepares the text data for this hybrid model, enhancing its ability to detect depressive sentiment. The algorithm's purpose, therefore, is to offer a dependable, expandable foundation for detecting depression, contributing to early identification and support in mental health monitoring efforts.

Algorithm 1: Hybrid deep learning-based depression detection (HDL-DD).

Algorithm: Hybrid Deep Learning-based Depression Detection (HDL-DD)
Input: Social media dataset D (Twitter)
Output: Results of depressive behavior R, performance statistics P

1) Begin
2) D'←DataPreprocessing(D)
3) selectedFeatures←FeatureEngineering(D')
4) embeddingSpace←GloVe(features)
5) (T1, T2)←DataPreparation(embeddingSpace, D')
6) Configure hybrid DL (DDNet) model m (as in Figure 2)
7) Compile m
8) m'←TrainDDNet(m, T1)
9) Save m'
10) Load m'
11) R←DepressionDetection(m', T2)
12) P←FindPerformance(real labels, R) //where R is predicted labels
13) Display R
14) Display P
15) End

To find sad behavior in social media data, the HDL-DD algorithm (Algorithm 1) uses tweets from Twitter. This method employs a hybrid DL model, feature engineering, and data preprocessing to identify melancholy accurately. The goal is to produce results that exhibit depressed behavior and assess the model's effectiveness. The technique begins by using preprocessing techniques to clean and normalize the raw dataset (D), resulting in a cleaned and normalized dataset (D'). Format standardization, noise removal (such as extraneous material or unusual characters), and duplicate entry removal are examples of preprocessing. By confirming the accuracy and consistency of the information, this stage increases the precision of the analysis that comes after. Following data cleansing, feature engineering is employed to find relevant traits indicating depressive language. To retain the most instructional portions of the data, these characteristics (selected characteristics) are carefully picked for sentiment analysis and depression diagnosis.

An encoding space with the selected properties is created using already trained GloVe word insertions. These embeddings convert to record the semantic connections between terms, creating a dense vector out of every word in the tweet representation. By mapping each word to a high- dimensional vector while maintaining contextual information, the embedding space enables the DL model to spot minute correlations and patterns in the text. The tweets are then categorized as word embedding sequences in this embedding space, leading to the development of the instruction and testing datasets (T1 and T2, respectively). Using this structure, the DL network can read each tweet as a set of pertinent word vectors. The method's foundation is configuring and training a hybrid DL model, or DDNet. CNN and LSTM networks are the building blocks of DDNet, as seen in **Figure 2**. Word pairings that can imply a negative attitude or gloomy tendencies are examples of local feature patterns that the CNN layers detect. However, the LSTM layers save information throughout the sequence, enabling the model to comprehend textual contextual links. Because it can identify sequential and localized language patterns linked to poor performance, this hybrid technique is especially well-suited for diagnosing depression.

When the model design is finished, the algorithm compiles (m), preparing it for training by configuring the proper optimizer, loss function, and performance measures. After that, the model is constructed as (m) and trained with the given dataset (T1). The model modifies its settings during training to increase detection accuracy by learning from patterns in the data that correlate with depressing language. After training, the learned model (m') is kept for future usage, allowing effective repurposing without retraining. Following training and saving, the model is loaded and applied to the test data (T2). Using the test set to detect melancholy, the loaded model (m') generates predictions (R), the algorithm's evaluations of depressive behavior in each tweet. A comparison between these predicted labels (R) and the actual labels is then used to evaluate the precision of the model. The model's capabilities to identify depressive behavior are quantitatively assessed by the performance evaluation, which includes measures such as accuracy, precision, recall, and $F1$-score, which are represented by (P). Finally, the algorithm displays the results (R) of the depression detection analysis and the performance statistics (P), giving insights into the detection outcomes and the model's accuracy. This structured approach allows the HDL-DD algorithm to process social media data effectively, leveraging hybrid deep learning to capture complex patterns of depressive language in social media posts, ultimately aiding in identifying potential depressive behavior in users.

3.5 Dataset Details

In this research, data was collected from a publicly available Twitter dataset (Sentiment140 Dataset) and filtered using depression-related keywords to extract relevant content. This filtering process resulted in a dataset of approximately 10,000 tweets. After preprocessing and deduplication, a total of 8,500 meaningful tweets were retained. These were then split into training and testing sets, with 6,800 tweets used for training and 1,700 for testing.

3.6 Evaluation Methodology

Since we employed measurements obtained from the confusion matrix, supervised learning, a learning-based technique shown in **Figure 4**, is used to evaluate our methodology. Since this is a binary classification problem, the model outputs a probability score for each tweet being depressive or non-depressive, which is then thresholded to make final predictions.

Effectiveness statistics are derived from a comparison between the reality of truth and the predicted labels of our method using the conflation matrix. Equations (1) through Equation (4) express the measures utilized in the performance evaluation.

$$Precision\left(p\right)=\frac{TP}{TP+FP} \tag{8}$$

$$Recall\left(r\right)=\frac{TP}{TP+FN} \tag{9}$$

$$F1-score = \frac{2*(p*r)}{(p+r)} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

The outcome of the measures for assessing achievement is a value between 0 and 1. ML research extensively uses these metrics. The evaluation measures given in Equation (8) to Equation (11) give us the opportunity to completely evaluate our model. For depression detection, having high recall means the model is able to catch most of the depressive tweets, which is crucial in order not to miss at-risk users. High precision means any flagged tweets are actually depressive, without too many false alarms. The claimed $F1$-score tries to take both into account, showcasing that the model can be stable. The $F1$- score of DDNet was 95%, which indicates the bodies' ability to learn the contextual patterns of depressive contents.

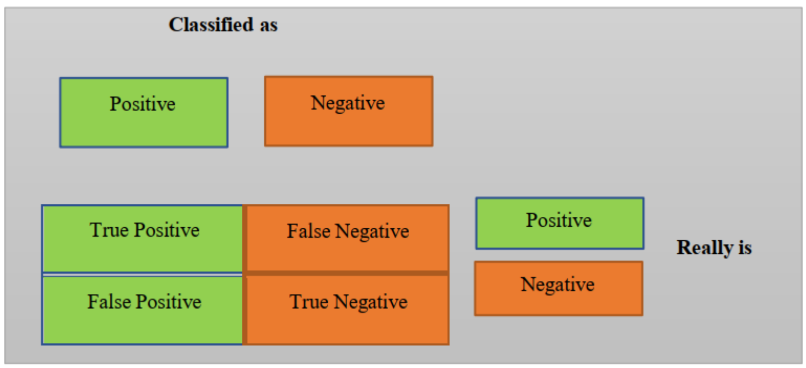

Figure 4. Confusion matrix.

4. Experimental Results

Here, we update our collaboration with a hybrid DL model that automatically identifies depression in social media messages. Our framework is implemented using the TensorFlow and Keras frameworks in Python 3. For our investigation, we used a dataset from Twitter. There is much promise for both academic and real-world mental health applications when it comes to identifying depression in Twitter tweets. Twitter offers a wealth of data that can be investigated to identify psychological trends and patterns. Using tweets to identify depression can help develop individualized mental health treatments depending on each person's needs and expressions.

Figure 5 shows a word cloud created using randomly selected tweets, highlighting the words that appear most frequently, including “thank,” “love,” “good,” “day,” and “hope.” Larger words show higher word frequency. Although such words are generally loaded with positive or neutral valences, and they inform rather than depress the identification of favourable terms, the tonal information in the detection task may be separated from the identification of favourable terms.

Figure 6 shows a word cloud generated from tweets that have been labeled as depressive, which includes emotionally weighted words as ‘depression’, ‘suicide’, ‘hopeless’, ‘lonely’, and ‘anxiety’. The use of terms such as “wide smile” also indicates concealed expressions of sadness. These frequent words indicate negative affective states, indicating that the model can achieve the identification of depressive cues in social media conversations.

Figure 7 shows the evolution of training and validation accuracy for the baseline RNN model through 20 epochs. The training accuracy approaches 91.43% at the end of the last epoch, and the validation accuracy is nearly 90.1%. Almost parallel curves in the plot contributed to consistent learning, while good convergence and overfitting were also observed, demonstrating the model's ability to generalize.

Figure 5. Word cloud generated using random tweets.

Figure 6. Word cloud generated using depressed tweets.

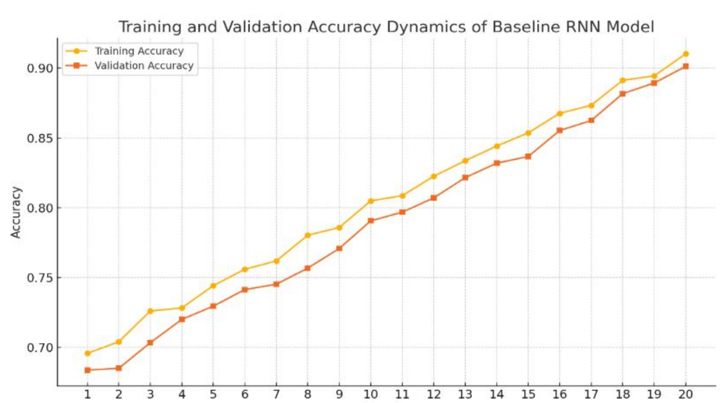

Figure 7. Epoch-wise training and validation accuracy curve of baseline RNN model.

Figure 8 shows a typical training and validation loss history for the baseline RNN model over 20 epochs. The training loss declines from about 0.6 to 0.18 while the validation loss falls to around 0.23. Decreasing trends observed in these curves suggest that learning for the models is reasonable and without much overfitting, and the model can be generalized to new datasets.

The training and validation accuracy of the baseline LSTM model over 20 epochs is illustrated in **Figure 9**. As the training accuracy rises, it plateaus at around 92.31%, while the validation accuracy almost catches up, reaching approximately 91%. The small value of the gap between the two curves suggests that the generalization is excellent, the learning is effective, and the LSTM outperforms its RNN baseline configuration.

The training and validation loss of the baseline LSTM model over 20 epochs is shown in **Figure 10**. The training loss decreases gradually to 0.18, while the validation loss remains at 0.21. The highly correlated and monotonically decreasing curves between these different numbers of training epochs suggest that the two RUL estimates are steadily converging towards each other, indicating the strong learning capability and excellent generalization of the LSTM model, with minimal overfitting.

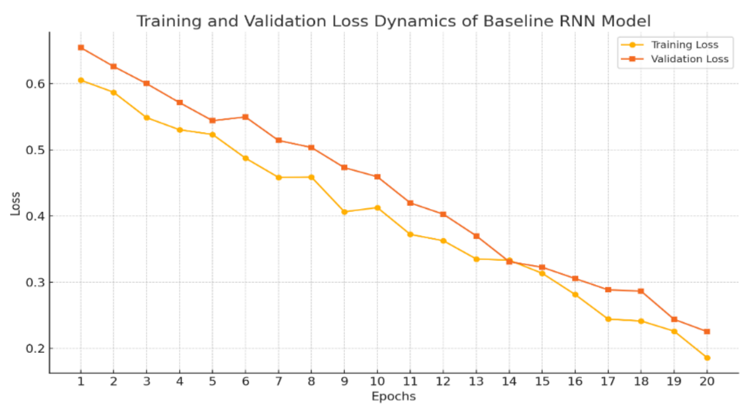

Figure 8. Epoch-wise training and validation loss curve of baseline RNN model.

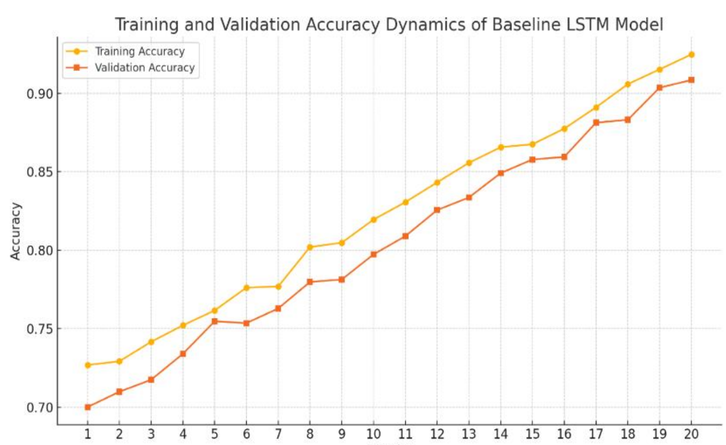

Figure 9. Epoch-wise training and validation accuracy curve of the baseline LSTM model.

Figure 11 provides a summary of the training and validation accuracy of the proposed hybrid model in 20 epochs. The training accuracy continues to increase, with a best training accuracy of 96.21%, and a validation accuracy of approximately 95%. The close matching and upward trends of both curves reveal that effective learning is taking place for our models, which exhibit strong generalization, and both yield better results than the baseline RNN and LSTM models.

Figure 12 presents the epoch-wise training and validation loss curves for the hybrid model with the proposed architecture trained for 20 epochs. The training loss drops from roughly 0.5 to 0.15, while the validation loss drops to 0.16. The similar downward trends and less space between the two curves indicate effective convergence, where we are not overfitting, and a better learning ability of the model.

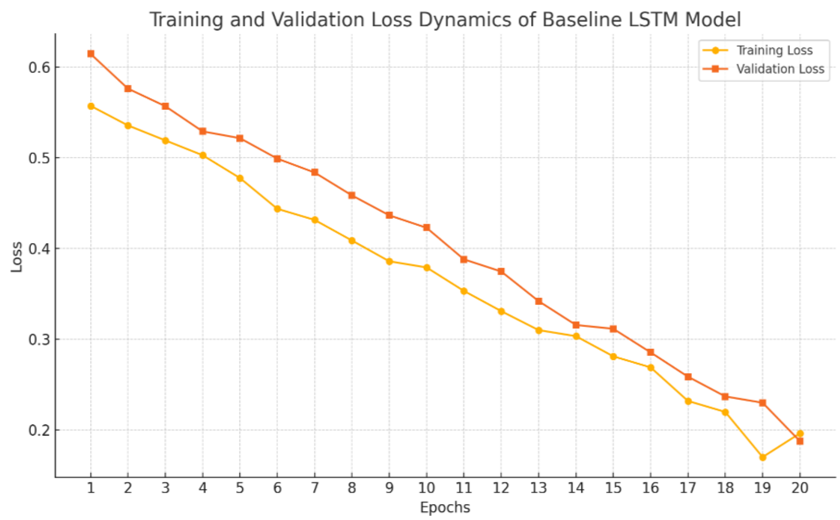

Figure 10. Epoch-wise training and validation loss curve of baseline LSTM model.

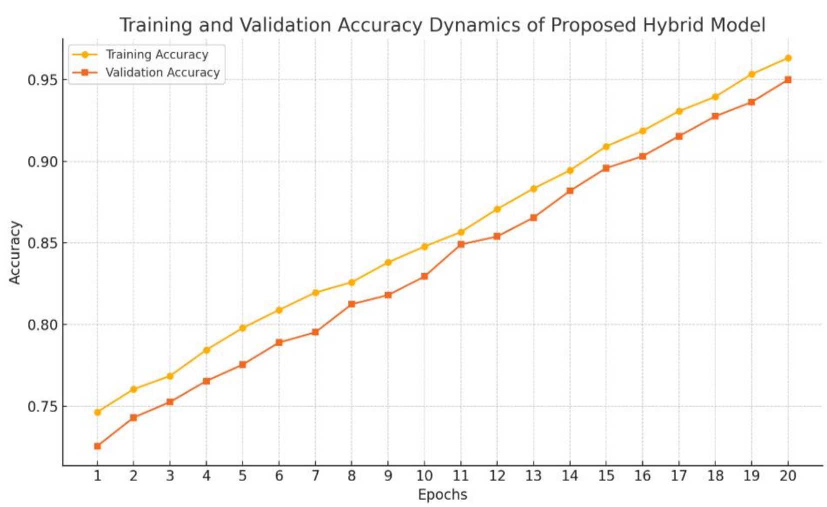

Figure 11. Epoch-wise training and validation accuracy curve of the proposed hybrid model.

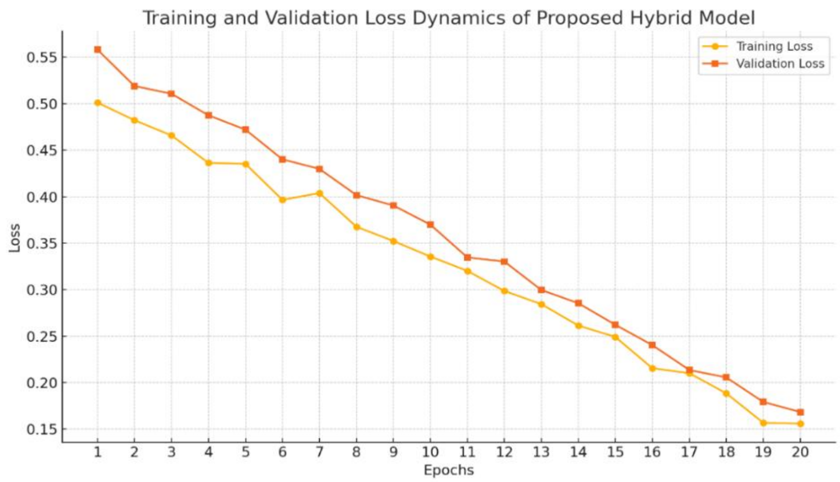

Figure 12. Epoch-wise training and validation loss curve of the proposed hybrid model.

Table 1 shows the comparative performance of the three models Baseline_RNN, Baseline_LSTM and the proposed Hybrid Model for depression detection. The maximum average scores are obtained by Hybrid Model proving its better performance with precision, recall, $F1$-score of 0.95, 0.96, 0.95 and accuracy of 96.21% which are better than the baselines in all the metrics, indicating its superior efficacy and dependability.

Table 1. Performance comparison among depression detection models.

Depression detection model	Precision	Recall	F1-Score	Accuracy
Baseline RNN	0.92	0.9	0.9	0.9143
Baseline LSTM	0.93	0.91	0.91	0.9231
Hybrid Model (Proposed)	0.95	0.96	0.95	0.9621

Figure 13 shows four sets of bar plots, i.e., (a) precision, (b) recall, (c) $F1$-score, and (d) accuracy, for Baseline RNN, Baseline LSTM, and the Hybrid Model, compared with the other. In subplot (a), precision increases from 0.92 in the RNN to 0.93 in the LSTM and reaches the maximum value of 0.95 in the Hybrid Model. Higher is better, which means a smaller number of false positive cases and more reliable results while identifying positive cases of depression. As shown in subplot (b), the recall increases from 0.90 for the RNN to 0.91 for the LSTM and then reaches 0.96 for the Hybrid Model. This significant improvement demonstrates the heightened sensitivity of the model in recognizing actual cases of depression.

Subplot (c) shows the $F1$-score (the harmonic mean of precision and recall). The RNN and LSTM achieve similar $F1$-scores of 0.90 and 0.91, respectively, while the hybrid model yields an $F1$-score of 0.95, indicating a stronger balance between positive detection and false detection. Subplot (d) presents the global accuracy, which increases from 91.43% with RNN to 92.31% with LSTM, and reaches 96.21% for the Hybrid Model. This demonstrates the excellent generalization ability of our model.

In general, the performance of the Hybrid Model is consistently superior to that of the RNN and LSTM baseline models across all four metrics, indicating that the Hybrid Model is effective and has excellent potential for robustness in detecting depression.

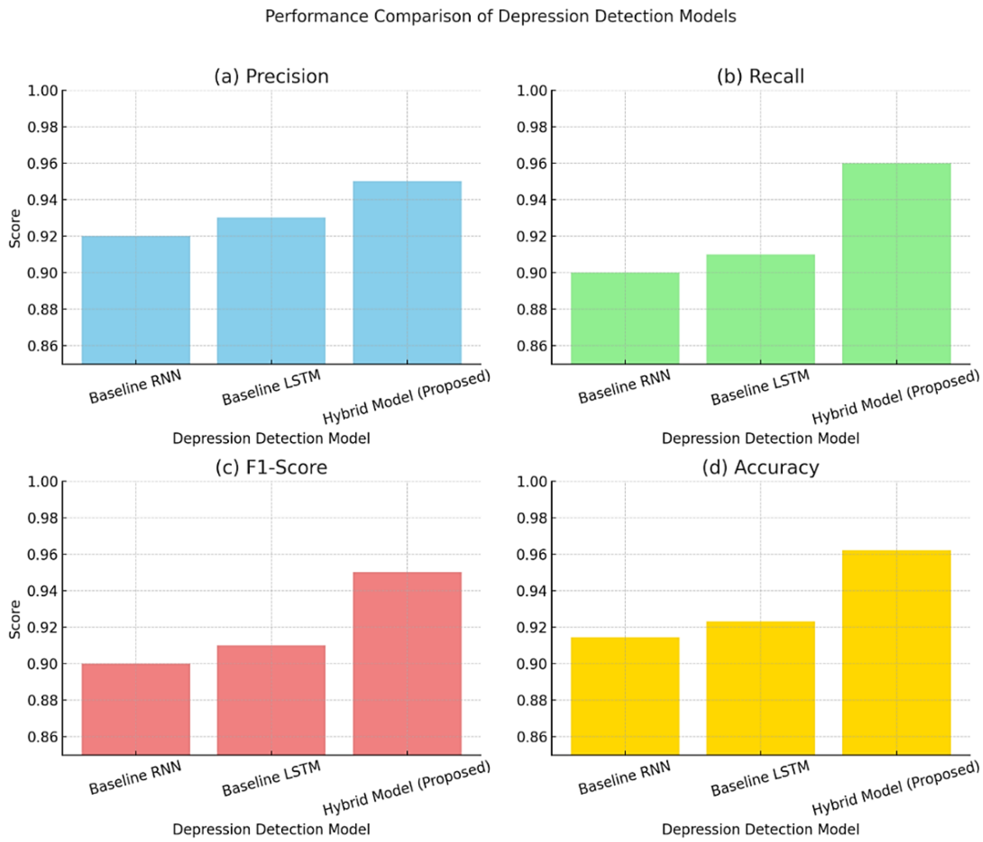

Figure 13. Comparative bar plots of precision, recall, $F1$-score, and accuracy for depression detection models.

Table 2 presents a performance comparison between the proposed DDNet model and existing depression detection models from the literature. The table highlights key metrics across different datasets: accuracy, precision, recall, and $F1$-score. Traditional machine learning models (SVM, RF) show lower accuracy, while deep learning-based approaches such as Hyper-graph-based Attention Models and Deep Integrated SVMs perform better. However, the proposed DDNet model (CNN- LSTM hybrid) outperforms all baseline models, achieving 96.21% accuracy and demonstrating its effectiveness in detecting depression from social media conversations. This superior performance validates the impact of hybrid architectures in mental health AI applications.

Table 2. Performance comparison of the proposed DDNet model with existing depression detection models.

Study	**Model**	**Dataset**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	***F*1-score (%)**
Alghamdi et al. (2020)	Machine Learning (SVM, RF)	Arabic Psychological Forum	85.40	83.2	84.7	83.9
Ahmed et al. (2022)	Hyper-graph- based Attention Model	Mental Health Text Corpus	91.50	90.1	91.0	90.5
Ding et al. (2020)	Deep Integrated SVM	College Student Depression Data	89.30	88.2	88.9	88.5
Cascarano et al. (2023)	Deep Learning + Longitudinal Analysis	Biomedical & Social Network Data	94.10	93.5	93.9	93.7
Proposed DDNet (This Study)	Hybrid CNN- LSTM	Social Media Depression Dataset	**96.21**	**95.0**	**96.0**	**95**

Figure 14 illustrates a performance comparison among various depression detection models, highlighting key metrics such as accuracy, precision, recall, and $F1$-score. Traditional machine learning models like SVM and RF (Alghamdi et al., 2020) show lower performance due to their reliance on handcrafted features, which fail to capture the complexity of depressive expressions in social media conversations. Deep learning models, such as Hyper-graph Attention (Ahmed et al., 2022) and Deep Integrated SVM (Ding et al., 2020), perform better by leveraging learned representations. Still, they struggle with sequential dependencies and context preservation. In contrast, the proposed DDNet (Hybrid CNN-LSTM model) achieves 96.21% accuracy, 95% precision, 96% recall, and 95% F1- score, surpassing existing models in detecting depressive content effectively.

DDNet's superior performance is attributed to its hybrid architecture, where CNN extracts local text features at the same time, LSTM retains long-term dependencies, allowing the model to analyze depressive expressions with higher contextual understanding. Unlike CNN-only models focusing on spatial features or RNNs suffering from vanishing gradients, DDNet effectively combines feature extraction and sequential learning for enhanced classification accuracy. Additionally, integrating pre- trained GloVe word embeddings improves semantic understanding, allowing the model to detect subtle depressive cues beyond explicit keywords.

Fine-tuned hyperparameters, such as an optimized learning rate (0.001), CNN filters (64), LSTM units (300), and dropout rate (0.5), contribute to the model's robust generalization and stability. Using dropout layers and early stopping further prevents overfitting, ensuring that DDNet can adapt to varied linguistic styles and text structures. Unlike prior models, which struggle to differentiate casual negative expressions from clinical depressive tendencies, DDNet effectively captures the gradual evolution of depressive discourse, making it a state-of-the-art AI-driven approach for depression detection.

The better performance of DDNet in **Figure 14**, and in particular in $F1$-score and Recall, indicates its advantage in recognizing depressive content with high coverage and accuracy. Our model performs better than all existing methods for real-time depression detection on all evaluation metrics, which satisfies the effectiveness of the CNN-LSTM hybrid architecture on real-time depression detection.

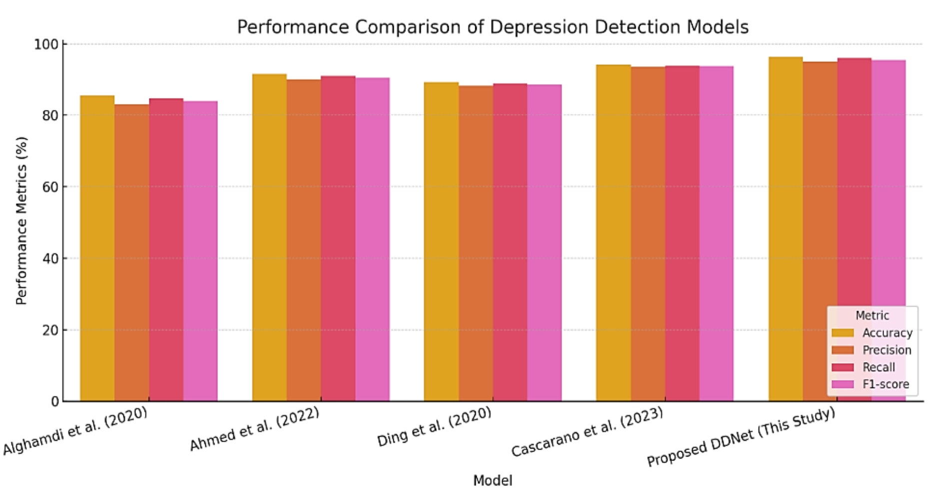

Figure 14. Performance comparison of depression detection models based on accuracy, precision, recall, and $F1$-score.

5. Discussion

The methodology using DL and social media messages can be automatically identified as sad and is presented in this study. Our literature analysis revealed that contextual information is frequently extracted from tweets utilizing CNN-style deep learning models. However, when it comes to data analysis for depression detection, LSTM models are better suited. We propose a hybrid DL model that combines CNN and LSTM to improve detection performance. We employ preprocessing and word embeddings to prepare the data for supervised learning. We collected and annotated our dataset from Twitter, a well-known social media platform.

Several tools and frameworks, such as TensorFlow, Keras, and Scikit-learn, were used for model development, training, and performance evaluation to perform the experimental assessment of DDNet. We used the Twitter API to collect the data and NLTK and spaCy for preprocessing processes, such as text tokenization, removing stopwords, and lemmatization. Pretrained word embeddings (GloVe) were then used for feature extraction from social media conversations, improving the model's performance in capturing context.

To evaluate the performance of the proposed model, we have compared the performance of DDNet with state-of-the-art deep learning models such as Baseline RNN, LSTM, and CNN architectures. As can be seen from Table 1, the proposed hybrid model also showed superior performance compared to traditional models with 96.21% accuracy, 95% precision, 96% recall, and 95% F1-score which is significantly better than baseline RNN (91.43%) and standalone LSTM (92.31%) for depression detection accuracy. They were highlighted as the key factor in improving the classification performance of CNN-LSTM hybrid models, which process the sequential segments of the input associated with the imbalanced dataset.

Our method achieves improved accuracy relative to state-of-the art techniques. For instance, Shah et al. (2020) reported an F1-score of 91.3% using a deep learning-based depression detection model, whereas Vandana et al. (2023) adopted a CNN-BiLSTM hybrid method, obtaining an accuracy of 93.7%. Additionally, it performs better than these benchmarks, yet it improves on the feature extraction and also interpretability as the final output from the model is more robust and directly applicable to real-time data. Unlike previous approaches that used lexicon-based sentiment analysis or stand-alone deep learning models, DDNet uses both CNN and LSTM components, learning localized word-based features and longer-term dependencies on both structured and unstructured text data. Our hybrid CNN-LSTM model performs significantly better than the standard CNN model for detecting depression in social conversations.

The superior performance of DDNet may be due to the architectural schema, which can represent contextual features deeper with stacked LSTM layers as well as suitable preprocessing for informal social media text. In contrast to classical systems based on shallow features or n-gram statistics, DDNet reflects linguistically more subtle cues such as hesitation, sarcasm, and the diffusion of negative sentiment. This is particularly useful in identifying early onset of depressive signs that might not be caught by conventional classifiers. Furthermore, SVM and RF don’t have sequential modeling and they are faster in the inference, which also makes their performance much lower. The model’s generalization is clear, as indicated by the close alignment of training and validation accuracy, but false positives (especially those with neutral-tone depressive posts) indicate that multimodal (e.g., image or user behavior patterns) should be integrated in future work.

Future research directions include extending multilingual datasets, real-time deployment, and integration with explainability techniques to enhance model interpretability and usability in clinical or mental health monitoring contexts. Results from experiments show that our hybrid deep learning model performs better than the state-of-the-art methods. However, some limitations are discussed in Section 5.1.

5.1 Limitations and Potential Failure Cases

While DDNet performs well in the task of detecting depression, some shortcomings have to be recognized. One of the main challenges is multi-lingual data, since it has been trained mainly on English text. Break-off depressions have different expressions in different languages and cultures, and direct cross-translation for such terms is often devoid of sentiment nuance. Any future enhancements can be targeted towards fine-tuning multilingual datasets, as well as making mBERT or other transformer-based models implemented for added learnability. Detection of mild forms of depression is another limitation. The model does well identifying explicit statements about depression but struggles to recognize subtle indicators like changes in writing tone, sarcasm, and marginally better or worse emotional placement. For this purpose, it may be helpful to enrich DDNet with contextual learning and user behavior analysis, improving its capacity to detect early-stage depression.

A potential problem is misclassification, where the tool may incorrectly flag non-depressive but pessimistic statements. Conversely, individuals who hide their feelings might go unnoticed. Incorporating multimodal data, such as voice tone and facial expressions, could help reduce these errors. Sarcasm and humor in social media noise make accurate classification more challenging. Ethical issues such as privacy and data security must also be considered. Future work will focus on privacy-preserving AI mechanisms that ensure responsible use of automated depression detection systems, aiming to promote transparency and trust among end users of such systems.

6. Social and Practical Implications

The increasing use of social media platforms realizes the original population medium on a self- expression and usually are a good source on catching down the mental health patterns! Considering the growing concern over mental health disorders, an AI approach like DDNet might prove not only practical but also social as well.

Social Implications

(i) *Early Detection and Intervention:* The DDNet framework suggests a system of identifying depression among social media users at the very first stage, and the potential automation with it can lead to providing psychological intervention at the right point in time. That is, when a person feels the desire to self-harm, they will be able to recognize that desire and seek professional help in advance, thus helping to reduce the possibility of developing a sharp mental health crisis.

(ii) *Normalize Mental Health:* With AI being integrated into mental health monitoring, the stigma behind discussing a topic like depression can be decreased.

Practical Implications

Integration into Mental Health Applications: The DDNet model could be integrated into mental health support systems, chatbots, or mobile applications to enable real-time depression detection and automated recommendations for mental health resources.

(i) *Work with health professionals:* It can support routine diagnosis and work with Psychologists and Psychiatrists by analysing social media behaviour. Tackling Corporate and Educational Well-being Programs: Enterprises and universities can utilize this AI-based framework to assess the mental well-being of the working class and pupils, thus creating customizable mental health support programs.

(ii) *Policy Makers and Government Implementation:* Policymakers could leverage AI-based depression detection insights to develop mental health awareness and implement data-driven public health policies.

(iii) DDNet is well- positioned to advance the field of mental wellness monitoring and interventions globally by ensuring scalability, reliability, and ethical data treatment. We will elaborate on that in upcoming research, in addition to improving the interpreter, a model that extends datasets to multilingual contexts, and improving the application of AI psychological tests.

7. Conclusion and Future Work

The proposed model DDNet is a hybrid deep learning architecture, which performs the detection and classification of depression for conversation data from social media using CNN and LSTM. The experimental results showed that DDNet outperforms the previous generation models in the same dataset, with an accuracy of 96.21%, compared to an LSTM (92.31%) and RNN (91.43%) standalone model. Our model is highly reliable in depression detection because it incorporates CNN's ability to learn local text patterns and LSTM's ability to learn long-range dependencies. Besides being highly accurate, our method provides a more scalable and adjustable deep learning model applicable to mental health apps, real-time social media monitoring, and clinical decision-support systems. Although our results are promising, our model has certain limitations that should be addressed in future research. First, the data used in this study are primarily from English-language tweets, which restricts the generalization of the model to multilingual and cross-cultural social media conversations. In summary, this study presents a new and practically applicable pipeline for automatic depression detection from social media contents. Using the complementary of CNN (strong in feature learning) and LSTM (strong in sequence memorization), the hybrid model DDNet outperforms previous models in terms of accuracy and context understanding. The model's excellent empirical performance and its applicability to real-time applications underlie its promise for large-scale digital mental health implementation. This work not only brings significant progress of deep learning techniques in affective computing, but also paves a new way for highly-ethical and theory interpretable AI in mental healthcare. Future work is advisable to collect multi-linguistic data and build multi-lingual depression detection models. Second, although DDNet can extract text features, it does not utilize multimodal data sources such as image, audio, or video data, which can also help to improve the accuracy of depression detection. This paper opens pathways for future research involving multimodal fusion methods to evaluate text, voice tone, and facial expressions together for a better mental health monitoring experience. Furthermore, AI-based depression detection models should be carefully examined regarding their ethical implications, such as considerations about privacy, potential biases of these models, or risks related to false positives.

Future work should address XAI approaches to enhance the interpretability of the model to produce transparent and trustworthy decision-making results in clinical and social contexts. Moreover, the real-time deployment of DDNet into mental health surveillance systems can lead to the integration of DDNet with chatbots, telemedicine portals, and mobile health applications, which can improve mental health intervention strategies. However, by tackling these challenges, AI-driven depression detection models such as DDNet can significantly contribute to improving mental health research, enabling earlier interventions, and promoting mental well-being in society.

Conflicts of Interest
The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria, educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgments
None.

AI Disclosure
The author(s) declare that no assistance is taken from generative AI to write this article.

References

Ahmed, U., Lin, J.C.W., & Srivastava, G. (2022). Hyper-graph-based attention curriculum learning using a lexical algorithm for mental health. *Pattern Recognition Letters*, *157*, 135-143. https://doi.org/10.1016/j.patrec.2022.03.018.

Alghamdi, N.S., Mahmoud, H.A.H., Abraham, A., Alanazi, S.A., & García-Hernández, L. (2020). Predicting depression symptoms in an Arabic psychological forum. *IEEE Access*, *8*, 57317-57334. https://doi.org/10.1109/access.2020.2981834.

Al-Omari, H., Abdullah, M.A., & Shaikh, S. (2020). Emodet2: Emotion detection in English textual dialogue using Bert and Bilstm models. In *2020 11th International Conference on Information and Communication Systems* (pp. 226-232). IEEE. Irbid, Jordan. https://doi.org/10.1109/icics49469.2020.239539.

Alsubhi, S.M., Alhothali, A.M., & AlMansour, A.A. (2023). AraBig5: the big five personality traits prediction using machine learning algorithm on Arabic tweets. *IEEE Access*, *11*, 112526-112534. https://doi.org/10.1109/ACCESS.2023.3297981.

Amin, S., Alharbi, A., Uddin, M.I., & Alyami, H. (2022). Adapting recurrent neural networks for classifying public discourse on COVID-19 symptoms in Twitter content. *Soft Computing*, *26*(20), 11077-11089. https://doi.org/10.1007/s00500-022-07405-0.

Amjad, M., Ashraf, N., Zhila, A., Sidorov, G., Zubiaga, A., & Gelbukh, A. (2021). Threatening language detection and target identification in Urdu tweets. *IEEE Access*, *9*, 128302-128313. https://doi.org/10.1109/ACCESS.2021.3112500.

Cao, L., Zhang, H., & Feng, L. (2020). Building and using personal knowledge graph to improve suicidal ideation detection on social media. *IEEE Transactions on Multimedia*, *24*, 87-102. https://doi.org/10.1109/tmm.2020.3046867.

Cascarano, A., Mur-Petit, J., Hernandez-Gonzalez, J., Camacho, M., de Toro Eadie, N., Gkontra, P., Chadeau-Hyam, M., Vitria, J., & Lekadir, K. (2023). Machine and deep learning for longitudinal biomedical data: a review of methods and applications. *Artificial Intelligence Review*, *56*(Suppl 2), 1711-1771. https://doi.org/10.1007/s10462-023-10561-w.

Ding, Y., Chen, X., Fu, Q., & Zhong, S. (2020). A depression recognition method for college students using deep integrated support vector algorithm. *IEEE Access*, *8*, 75616-75629. https://doi.org/10.1109/access.2020.2987523.

Elsafoury, F., Katsigiannis, S., Pervez, Z., & Ramzan, N. (2021). When the timeline meets the pipeline: a survey on automated cyberbullying detection. *IEEE Access, 9*, 103541-103563. https://doi.org/10.1109/access.2021.3098979.

Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering, 30*(3), 1819-1842. https://doi.org/10.1007/s11831-022-09863-z.

Gedam, S., & Paul, S. (2021). A review on mental stress detection using wearable sensors and machine learning techniques. *IEEE Access, 9*, 84045-84066. https://doi.org/10.1109/access.2021.3085502.

Grover, P., Kar, A.K., & Dwivedi, Y.K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research, 308*(1), 177-213. https://doi.org/10.1007/s10479-020-03683-9.

Huang, G., Li, Y., Jameel, S., Long, Y., & Papanastasiou, G. (2024a). From explainable to interpretable deep learning for natural language processing in healthcare: How far from reality?. *Computational and Structural Biotechnology Journal*, *24*, 362-373. https://doi.org/10.1016/j.csbj.2024.05.004.

Huang, Y., Guo, J., Chen, W.H., Lin, H.Y., Tang, H., Wang, F., & Bian, J. (2024b). A scoping review of fair machine learning techniques when using real-world data. *Journal of Biomedical Informatics*, *151*, 104622. https://doi.org/10.1016/j.jbi.2024.104622.

Khan, S., Kamal, A., Fazil, M., Alshara, M.A., Sejwal, V.K., Alotaibi, R.M., Baig, A.R., & Alqahtani, S. (2022). HCovBi-caps: hate speech detection using convolutional and Bi-directional gated recurrent unit with Capsule network. *IEEE Access*, *10*, 7881-7894. https://doi.org/10.1109/ACCESS.2022.3143799.

Kumar, V., Recupero, D.R., Riboni, D., & Helaoui, R. (2021). Ensembling classical machine learning and deep learning approaches for morbidity identification from clinical notes. *IEEE Access, 9*, 7107-7126. https://doi.org/10.1109/access.2020.3043221.

Li, D., Rzepka, R., Ptaszynski, M., & Araki, K. (2020a). HEMOS: a novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Information Processing & Management*, *57*(6), 102290. https://doi.org/10.1016/j.ipm.2020.102290.

Li, L., Ma, Z., & Cao, T. (2020b). Leveraging social media data to study the community resilience of New York City to 2019 power outage. *International Journal of Disaster Risk Reduction*, *51*, 101776. https://doi.org/10.1016/j.ijdrr.2020.101776.

Murshed, B.A.H., Abawajy, J., Mallappa, S., Saif, M.A.N., & Al-Ariki, H.D.E. (2022). DEA- RNN: a hybrid deep learning approach for cyberbullying detection in Twitter social media platform. *IEEE Access, 10*, 25857-25871. https://doi.org/10.1109/ACCESS.2022.3153675.

Nash, C., Nair, R., & Naqvi, S.M. (2023). Machine learning in ADHD and depression mental health diagnosis: a survey. *IEEE Access, 11*, 86297-86317. https://doi.org/10.1109/ACCESS.2023.3304236.

Rahman, R.A., Omar, K., Noah, S.A.M., Danuri, M.S.N.M., & Al-Garadi, M.A. (2020). Application of machine learning methods in mental health detection: a systematic review. *IEEE Access, 8*, 183952-183964. https://doi.org/10.1109/access.2020.3029154.

Rao, G., Zhang, Y., Zhang, L., Cong, Q., & Feng, Z. (2020). MGL-CNN: a hierarchical posts representations model for identifying depressed individuals in online forums. *IEEE Access, 8*, 32395-32403. https://doi.org/10.1109/access.2020.2973737

Safa, R., Bayat, P., & Moghtader, L. (2022). Automatic detection of depression symptoms in Twitter using multimodal analysis. *The Journal of Supercomputing*, *78*(4), 4709-4744. https://doi.org/10.1007/s11227-021-04040-8.

Salvi, M., Loh, H.W., Seoni, S., Barua, P.D., García, S., Molinari, F., & Acharya, U.R. (2024). Multi-modality approaches for medical support systems: a systematic review of the last decade. *Information Fusion*, *103*, 102134. https://doi.org/10.1016/j.inffus.2023.102134.

Sánchez, G.C., Marques, G., Dorronzoro, E., Romero, O.R., Martín, M.F., & De la Torre-Díez, I. (2020). Suicide risk assessment using machine learning and social networks: a scoping review. *Journal of Medical Systems, 44*(12), 1-15. https://doi.org/10.1007/s10916-020-01669- 5.

Sentiment140 dataset. Retrieved from https://www.kaggle.com/datasets/kazanova/sentiment140.

Shah, F.M., Ahmed, F., Joy, S.K.S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M.H. (2020). Early depression detection from social network using deep learning techniques. In *2020 IEEE Region 10 Symposium* (pp. 823-826). IEEE. Dhaka, Bangladesh. https://doi.org/10.1109/tensymp50017.2020.9231008.

Sujith, A.V.L.N., Sajja, G.S., Mahalakshmi, V., Nuhmani, S., & Prasanalakshmi, B. (2022). Systematic review of smart health monitoring using deep learning and Artificial intelligence. *Neuroscience Informatics*, *2*(3), 100028. https://doi.org/10.1016/j.neuri.2021.100028.

Thati, R.P., Dhadwal, A.S., Kumar, P.S. (2023). A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. *Multimedia Tools and Applications, 82*(4), 4787-4820. https://doi.org/10.1007/s11042-022-12315-2.

Uban, A.S., Chulvi, B., & Rosso, P. (2021). An emotion and cognitive-based analysis of mental health disorders from social media data. *Future Generation Computer Systems, 124*, 480-494. https://doi.org/10.1016/j.future.2021.05.032.

Uddin, M.Z., Dysthe, K.K., Følstad, A., & Brandtzaeg, P.B. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, *34*(1), 721-744. https://doi.org/10.1007/s00521-021-06426-4.

Vandana, Marriwala, N., Chaudhary, D. (2023). A hybrid model for depression detection using deep learning. *Measurement: Sensors*. *25*, 1-10. https://doi.org/10.1016/j.measen.2022.100587.

Wang, C., Wu, Y., Wang, C., Ren, Y., Shen, J., Pang, T., & Yu, Y. (2024). MSFNet: A multi-scale space-time frequency fusion network for motor imagery EEG classification. *IEEE Access*, *12*, 8325-8336. https://doi.org/10.1109/ACCESS.2024.3351204.

Wang, R., Bashyam, V., Yang, Z., Yu, F., Tassopoulou, V., Chintapalli, S.S., Skampardoni, I., Sreepada, L.P., Sahoo, D., Nikita, K., & Davatzikos, C. (2023). Applications of generative adversarial networks in neuroimaging and clinical neuroscience. *Neuroimage*, *269*, 119898. https://doi.org/10.1016/j.neuroimage.2023.119898.

Wang, T., Lu, K., Chow, K.P., & Zhu, Q. (2020). COVID-19 sensing: negative sentiment analysis on social media in China via BERT model. *IEEE Access*, *8*, 138162-138169. https://doi.org/10.1109/access.2020.3012595.

William, D., & Suhartono, D. (2021). Text-based depression detection on social media posts: a systematic literature review. *Procedia Computer Science, 179*, 582-589. https://doi.org/10.1016/j.procs.2021.01.043.

Xu, Q.A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: emerging trends and challenges. *Decision Analytics Journal*, *3*, 100073. https://doi.org/10.1016/j.dajour.2022.100073.

Yang, K., Zhang, T., & Ananiadou, S. (2022). A mental state knowledge–aware and contrastive network for early stress and depression detection on social media. *Information Processing & Management*, *59*(4), 102961. https://doi.org/10.1016/j.ipm.2022.102961.

Yue, C., Ware, S., Morillo, R., Lu, J., Shang, C., Bi, J., Kamath, J., & Wang, B. (2020). Automatic depression prediction using internet traffic characteristics on smartphones. *Smart Health*, *18*, 100137. https://doi.org/10.1016/j.smhl.2020.100137.

Zogan, H., Razzak, I., Wang, X., Jameel, S., & Xu, G. (2022). Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web, 25*(1), 281-304. https://doi.org/10.1007/s11280-021-00992-2.