DETECTING DEPRESSION IN SOCIAL MEDIA USERS USING ADVANCED DEEP LEARNING APPROACHES

Thesis Submitted
in Partial Fulfillment of the Requirements for the
Degree of

MASTER OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE

by

Abhishek Vipul Vora
(2K23/AFI/01)

Under the supervision of
Dr. Prashant Giridhar Shambharkar
Assistant Professor, Department of Computer Science & Engineering
Delhi Technological University

Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

MAY, 2025

CANDIDATE’S DECLARATION

I, **Abhishek Vipul Vora (23/AFI/01)**, hereby certify that the work which is being presented in the major project report II entitled “**Detecting Depression in Social Media Users using Advanced Deep Learning Approaches**” in partial fulfillment of the requirements for the award of the Degree of Master of Technology, submitted in the **Department of Computer Science and Engineering**, Delhi Technological University is an authentic record of my own work carried out during the period from **August 2023** to **May 2025** under the supervision of **Dr. Prashant Giridhar Shambharkar.**

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor (s) **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

I hereby certify that the Project titled “**Detecting Depression in Social Media Users using Advanced Deep Learning Approaches**”, submitted by **Abhishek Vipul Vora**, Roll No. 23/AFI/01, Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of **Master of Technology** (M.Tech) in Artificial Intelligence is a genuine record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in par or full for any Degree to this University or elsewhere.

Dr. Prashant Giridhar Shambharkar
Assistant Professor
Department of CSE
Delhi Technological University

Date:

ACKNOWLEDGEMENT

I am grateful to Prof. Manoj Kumar, HOD (Department of Computer Science and Engineering), Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, and all other faculty members of our department for their astute guidance, constant encouragement, and sincere support for this project work.

I am writing to express our profound gratitude and deep regard to my project mentor Dr Prashant Giridhar Shambharkar, for his exemplary guidance, valuable feedback, and constant encouragement throughout the project. His valuable suggestions were of immense help throughout the project work. His perspective criticism kept us working to make this project much better. Working under him was an extremely knowledgeable experience for us. I would also like to thank all my friends for their help and support sincerely.

Abhishek Vipul Vora

DETECTING DEPRESSION IN SOCIAL MEDIA USERS USING ADVANCED DEEP LEARNING APPROACHES

ABHISHEK VIPUL VORA

ABSTRACT

Due to stigma and a lack of easily available diagnostic resources, depression is a common mental health illness that frequently goes undiagnosed. As social media has grown in popularity, people are expressing their feelings and psychological states more and more online, which offer a wealth of information for mental health research. In order to detect depression from multimodal data—specifically, textual content from social media and user behavioral features—this thesis proposes a hybrid deep learning model that combines Bidirectional Encoder Representations from Transformers (BERT), Bidirectional Long Short-Term Memory (BiLSTM), and Convolutional Neural Networks (CNN).

The design makes use of CNN's prowess in identifying local patterns in text embedding, BiLSTM's capacity to model sequential relationships, and BERT's contextual language understanding to capture subtle semantic representations. To enhance the model's input space, behavioral data like posting frequency, activity time, and interaction patterns are combined with textual features. A carefully selected multimodal dataset was used to train and assess the hybrid model, which was then contrasted with a number of baseline models, such as deep learning variations and conventional machine learning classifiers.

In terms of accuracy, precision, recall, and AUC, experimental results show that the suggested hybrid model performs noticeably better than baseline methods, underscoring its efficacy and resilience in identifying depressed symptoms. This thesis offers a solid basis for further study in the area of intelligent psychological diagnosis and highlights the possibilities of integrating linguistic and behavioral modalities for automated mental health assessment.

Table of Content

List of Tables

List of Figures

List of Abbreviations, Symbols, and Nomenclature

WHO – World Health Organization

NLP – Natural Language Processing

BERT – Bidirectional Encoder Representation from Transformer

RNN – Recurrent Neural Network

CNN – Convolutional Neural Network

LSTM – Long Short Term Memory

BiLSTM – Bidirectional Long Short Term Memory

VADER – Valence Aware Dictionary and Sentiment Reasoner

GELU – Gaussian Error Linear Unit

MCC – Matthew's Correlation Coefficient

ROC – Receiver Operating Characteristic

AUC – Area Under Curve

CHAPTER 1
INTRODUCTION

1.1 Overview

The extensive use of various multiple social media platforms in recent years have changed the way how all the people interact with each other or communicate and express themselves. Millions of people now use these platforms as a medium like a diary that captures their feelings, ideas, there mood and even their mental health conditions. They can open their mind and heart out on these platforms and interact with different new users from all around the world. With more than 4.7 billion users worldwide, social media provides a large amount of text, behavioral, image data that when properly examined can provide insights into user's psychological health in addition to serving as a platform for social interaction with different people [1].

In the twenty first century, mental health in particular depression has become an important area of concern. One of the most common and prevalent mental health conditions is depression. It is characterized by persistent low mood, loss of interest or enjoyment in activities that once seemed pleasurable and cognitive impairment. One of the renowned organizations of the World Health Organization (WHO), gives an estimation that depression is one of the major cause of disability and affects more than 280 million people worldwide [2]. It is a matter of concern as a large number of people with depression choose not to take professional help because of stigma, ignorance or lack of services for mental health. However those individuals may unintentionally signal emotional distress through various online behaviors like the use of language, rate at which they are posting online, visual content shared by them and the way they are engaging with other users [3].

As a result of this insight, the multidisciplinary field of computational mental health has emerged, in which patterns which are linked to mental health issues can be found in digital footprints using various artificial intelligence techniques such as machine learning, deep learning and natural language processing techniques. In particular, deep learning, a branch of machine learning has proven to be incredibly effective in processing all the unstructured data such as text, photos and audio. As such it is a powerful method for social media data analysis.

This thesis addresses the potential of using advanced deep learning approaches to detect warning signs of depression among social media users in order to inform efforts to detect early and perhaps improve outcomes of mental health in patients by interventions as they timely.

1.2 Motivation

The motivation behind this study arises from the increasing prevalence of mental health disorders, alongside the enormity of the missing access to scalable, affordable and timely mental health diagnosis systems. The traditional modalities used to diagnose such cases are systems that require physical interviews and even some requires clinical assessments, which are resource intensive and are often unavailable, particularly in low resource or rural settings. In addition to this there are numerous people who are reluctant to seek help out of fear of stigma in society, as well as fear from their own side. Therefore, tens of millions of people go untreated and this increases the chance of the symptoms worsening and perhaps even committing suicide.

At the same time the digital age has created a new paradigm where people will willingly post intimate details about their lives online. Such statements, whether a tweet about feeling worthless, complaint about insomnia or a variation in tone and intensity of interaction, can be good indicators of some underlying mental health problem. Through research it can be seen that people suffering from depression are more likely to post more about negative emotions, use more first person pronouns (I, me etc.) and reduced social interaction.

The key motivation for this research is the capability to capture these more subtle digital cues by using the more sophisticated deep learning models, able to detect both linguistic and visual pattern which may otherwise be lost in traditional analysis [4]. In light of the emergence of transformative models such as Bidirectional Encoder Representation from Transformer BERT, there is a unique opportunity to develop a paradigm of robust, data driven systems for passive and continuous mental health monitoring [5].

Such systems can help psychologists, researchers, and the authorities that concern public health discover at-risk people early and prescribe them relevant intervention strategies. In addition, automated tools that are based on deep learning can scale to millions of users — and make them exceptionally valuable in our digitally connected, but mentally vulnerable society.

1.3 The Role of Social Media in Mental Health Detection

The social media platforms measure the user’s emotions and thoughts on a daily basis. Whether it’s in the form of a short tweet or an extended Reddit post, users tend to vent, request help or share personal hardships in a cyberspace. These phrases include linguistic, affective, and temporal clues that on a holistic analysis can reveal the behavioral signatures of depression. For example:

- **Linguistic patterns:** The depressed persons tend to use more first person pronouns (“I”, “me”) and negative sentiment words and lower linguistic diversity.
- **Temporal behavior:** Delayed night posts or sudden withdrawal or discontinue may be an indication.
- **Visual content:** Color selection, what’s chosen to be in focus and the use of filters in shared images might reveal how the poster feels.
- **Social engagement:** Impaired links with other people through internet, or unpredictable changes in virtual friendship can be associated with isolation.

These signals are very weak and everyone displays them in different ways which makes it impossible to manually measure them for large groups of people. As a result, this study will aim at developing automated scalable and intelligent systems capable of denoting insights from these cues to predict the possibility of depression cases using deep learning.

1.4 Deep Learning for Depression Detection

The area of deep learning has radically changed such areas like image recognition and natural language processing (NLP). It is one of its great strengths the capacity to recognize complex data patterns on its own eliminating dependence on expert-designed rules. Deep learning provides strong mechanisms for discovering complex and diverse signals embedded in the content provided by social media users for the purpose of depression detection. The analysis of this thesis is based upon the examination of three critical facets of user signals – text, time and behavior – to determine the complicated ways in which depression reveals itself within digital communication. Utilization of transformer-based models [including BERT] will allow one to understand the user text’s semantic context at a deeper level, allowing the interpretation not only of keywords but of the implicit meanings relating to the form of grammar, user objectives, and emotional tones [6]. Temporal modeling is very important in the observation of the development of depressive symptoms by means of using machine learning methods such as RNNs or LSTMs that are the appropriate methods for processing temporal sequences of user activity. Furthermore, to identify typical indicators of depressions, post frequency, time of activity, interaction metrics and alterations in language use are scrutinized [7].

The combination of these modalities results in a new hybrid deep learning system that combines textual context, temporal trends and behavioral signals for a more holistic and informed process for detecting depression in social media users. A sequence of tough experiments is carried out to examine the effectiveness of the proposed architecture, and the results are benchmarked against the performance metrics of regular deep learning approaches.

1.5 Objectives

This thesis aims to develop, improve, and test a potent hybrid deep learning architecture aimed at detecting depression symptoms in individuals on social media platforms. Leveraging an armory of social media inputs ranging from written expression and messaging rhythms to certainty-based observation proxies, this work aims to optimize virtual depression assessments for greater reliability, real-time service, and user self-analysis. This methodology departs from traditional keyword-based to one that is more complete, and one that can reveal psychological signals, subtle in social media activity, through the analysis of data.

The specific objectives of the thesis are:-

- Analyzing the posts of the users using the state of the art technologies, for example BERT in order to discover their sentiment, meaning and emotional character.
- To analyze time based user dynamics with models like RNN’s and LSTM’s to identify evolving changes with depressive symptoms.
- Including behavioral attributes as ways user’s mental health can be revealed.
- To build a hybrid deep learning architecture using the text, temporal and

behavioral features, improving the performance of the model for depression detection.

- To compare the performance of the proposed model with the conventional and traditional deep learning models by common standards.

1.6 Challenges

The task of detecting depression from social media data poses a variety of substantial challenges for deep learning. In the first place, it is hard to collect high quality labeled data, as mental health is a sensitive issue and clinical annotations are seldom provided in public collections. The data typically displays an imbalance, as posts indicating depression are outnumbered by those with neutral or positive emotions, while user activity patterns can also be inconsistent, both of which complicate attempts at temporal modeling [8]. Since emotions are expressed differently by users in different contexts and cultures, models struggle to consistently detect depression using linguistic signals. The inconsistent timing and fluctuating length of user sequences make temporal modeling with RNN and LSTM’s especially hard [9]. Although behavioral patterns may be useful but this patterns may not always represent mental state so well, due to their noisy nature. Besides the use of deep learning in mental health applications, it is limited by the lack of explanations from models, especially those combining different techniques. In addition to this, to protect user privacy and promote ethical practices, a great care needs to be taken while taking into consideration the social media information for mental health inference.

CHAPTER 2
RELATED WORK

2.1 Literature Survey

Deep learning methods for identifying and evaluating mental health issues using multimodal data from social media sites have advanced significantly in recent years. Researchers have created novel frameworks that integrate textual, visual, and emotive information to improve the accuracy of diagnosing depression. By enhancing model robustness, computing efficiency, and scalability, these innovations enhance proactive mental health therapies and allow for the early diagnosis of depressed symptoms. Let’s look at a quick rundown of how these technologies have been applied in studies to identify depression.

Lim et al. [10] introduced a new multi-modal system for forecasting depression risks using data from social networks. In this research, study data includes text and images from social media, rather than just text as is used in common studies. Using this early fusion method in transformers, the approach managed to surpass the multi-modal versions currently leading in the field. It points out that only a few posts are needed for effective predictions, so risks of depression can be found as soon as possible. This research uses time2vec for time-based representation and a CLIP model to process both text and picture information about posts. This research stresses that using several methods can make it easier to find depression on social media platforms.

A machine learning approach to find depression among social media users was introduced by Aik Seng Liaw et al. [11]. In order to develop their model, the researchers performed Twitter scraping on 1,295 accounts. To get an excellent F1-score of 82.05% and impressive accuracy, XGBoost was put to use in this model. The work demonstrates that by using machine learning and data from social media, accuracy in predicting depression is increased in AI-based mental health screening.

By analyzing data from social media, Amanat et al. [12] aimed to proven a method for detecting depression at its earliest stages. The data for this study was obtained by scraping Twitter and placed on Kaggle. A range of machine learning models, including CNN, SVM, NB, DT and their suggested RNN-LSTM design, was used by the authors to find out if a patient had depression. The results confirmed that the RNN-LSTM model had the best accuracy of all the methods, with 99.6%. After that, SVM, NB, CNN and DT placed third, with 97.21%, 97.31% and 82%, respectively. Researchers found that RNN-LSTM makes relatively few mistakes and is able to correctly identify depression cases. Further research might test hybrid types of RNN algorithms to see if they can improve early identification of depression.

A high-performing sentiment analysis model using double bidirectional gated recurrent units (BiGRU) was put forward by Han et al. in [13]. They brought together text files, pictorial emoticons and tags collected from SentiDrugs. I compared the suggested approach with LSTM, BiGRU, Memnet and Cabasse, among other models. Results showed that BiGRU gave the best results, with 78.6% accuracy.

The team of Liu et al. [14] developed a text classification approach based on the Medical Social Media Text Classification technique for screening consumer health. The chosen method improved the use of some consumer health terms in medical text classification, showing an accuracy of 87.65%.

Using details from pictures posted to Instagram, Reece et al. [15] relied on machine learning methods to discover signs of depression. Classification and strength analysis were done by using Random Forest (RF) and Bayesian Logistic Regression (LR). Using their model, they achieved strong results, getting an F1 of 64%, recall of 67.7% and precision of 64%.

Using SMPD (Social Media Prediction Dataset), Zhang and Wang et al. [16] have vastly enhanced the way user behavior and social media popularity are examined and predicted. To do so, they relied on measuring how people use social media experiments using multiple types of data and treated the prediction problem as one that needed to be done as events unfolded. Their methods were able to spot regular actions people follow after sharing information online. From evaluations using MAE

8 and SRC, their temporal modeling resulted in reliable and straightforward predictions useful for studying behavioral and mental health trends. They believed that predicting user activities on social media was urgent and used different sources to help with this. Thanks to their modeling process, they were able to see what kinds of patterns emerged in the way people behaved after sharing. Based on measures such as Spearman Ranking Correlation and MAE, it was found that the authors approach predicts trustworthy and clear results for analyzing behavior and mental health trends in people.

These studies collectively showcase the rapidly evolving landscape of deep learning applications in the field of mental health, specifically for detecting depression using social media data. As outlined in Table 2.1, each study offers distinct methodologies and approaches for predicting depression, highlighting the potential of combining textual, visual, and behavioral data from platforms like Twitter, Instagram, and Facebook. These works contribute valuable insights into the use of machine learning and deep learning techniques for early depression detection, emphasizing the promise of leveraging user-generated content for mental health analysis.

Table 2.1- Literature Survey of Recent Depression detection research papers.

Author	Year, Reference	Model Used	Dataset Used	Performance	Findings
Zong et al. [17]	2023	BiLSTM , BERT , BioBERT , MentalBERT , EAN , ERAN	eRisk2018	F1 score:- ECD = 0.59, BERT = 0.54, BioBERT = 0.52, ERAN = 0.47	The model achieves superior results compared to baseline models, as evidenced by higher precision, recall, and F1 scores.
Amanat et al. [12]	2022	Proposed RNN-LSTM approach, CNN, SVM.	Tweets-scraped dataset (Kaggle)	Accuracy:- CNN = 91% SVM = 97.21% RNN-LSTM = 99.6%	With a low error rate, the suggested RNN-LSTM approach produced the best results.
Liaw et al. [11]	2022	XGBoost model used.	Twitter Scraped dataset (1,295 Twitter users data)	F1- score of around 82.05%	With at least double the significance of other variables, the quantity of depression-related keywords in liked tweets turned out to be the most important one.

Wang et al. [18]	2020	Fine-tuned BERT model is proposed, SVM, CNN, LSTM are the othr models used.	Weibo	Accuracy: Proposed BERT model accuracy was around 75.6%. Other best performing model was of CNN with 71.1%.	The suggested improved model can be effectively applied to sentiment analysis.
					The study can be expanded in the future by utilizing more possible social media channels.
Ruz et al. [19]	2020	SVM, RF, TAN	Two Twitter datasets used. One was 2010 Chilean earthquake and the other 2017 Catalan independence referendum.	Accuracy: BF TAN and SVM are the two best performing models.	The SVM model is a reliable tool for analysing depression under adverse circumstances.
Han et al. [13]	2020	BiGRU, LSTM and multiple other models used	SentiDrugs	The proposed PM-DBiGRU model gave the highest accuracy of 78.6%	The suggested method can improve the effectiveness of drug level aspect-reviews.
Ahmad et al. [20]	2019	KNN, RF, and other deep learning methods like CNN, CNN+LSTM used	Twitter	KNN and RF gave around 72% and 82% accuracy. CNN and CNN+LSTM gave around 83% and 92% accuracy.	By storing both recent and historical data, the suggested model combined produced the best prediction results.

Reece et al. [15]	2017	Models used were Bayesian Logistic Regression and RF	Instagram + CES-D	Precision, Recall and F1 score were calculated as 64%, 67.7% and 64%	The proposed approach shown significant results outperforming other methods.
Lin et al. [21]	2014	SVM, RF, NB, DNN (proposed)	Weibo dataset was used.	DNN model gave the highest accuracy of 78.5% followed by RF giving 76.7% accuracy.	This model could help efficiently detect depression early signs and symptoms in patients.

CHAPTER 3
GENERALIZED FRAMEWORK

Here we are outlining a generalized model for depression detection. The various steps involved are:

1. Data Acquisition:

This is the first and crucial step where a dataset of tweets and images are gathered. The images and tweets will encompass a range of depressive and not depressive images and texts along with emoticons. The dataset's quality and diversity are crucial factors for the model's success.

2. Dataset Partitioning: Training Set, Validation Set, and Test Set:

The gathered data is broken into three distinct segments: the training set, validation set, and test set.

- The training set is employed to train the learning model, enabling it to identify patterns and characteristics that are indicative of the health or illness status in apple leaves.
- The validation set is utilized to optimize the model parameters and mitigate the risk of overfitting. This dataset facilitates model optimization by offering feedback on its performance on previously unseen data.
- The test set is employed subsequent to the completion of model training and fine-tuning. It offers an impartial assessment of the performance of the ultimate model, providing an estimation of its effectiveness on data that has not been previously observed.

3. Train the Model:

During this stage, the training data is inputted into a machine learning algorithm in order to acquire knowledge from it. This procedure entails feature extraction and feature selection, in which the algorithm discerns and acquires the most significant characteristics from the data that serve as indicators for different diseases.

4. Fine tune the Model:

After completing the initial training, the model's performance is evaluated using the validation set. According to this evaluation, the model's hyper parameters are modified to enhance its accuracy and ability to generalize. This is an iterative procedure that persists until the model achieves satisfactory performance on the validation data.

5. Evaluate the Model:

Subsequently, the model that has undergone training and fine-tuning is assessed using a collection of measures, which include accuracy, precision, recall, and F1 score, to ascertain its performance. This evaluation is based on how well the model predicts the correct categories on the validation set.

6. Predictive Model:

Once a predictive model has undergone training, fine-tuning, and evaluation, it is prepared to make predictions on fresh data. In this scenario, it will classify the tweets and images into two categories ‘Depressive’ or ‘Not Depressive’.

CHAPTER 4

PROPOSED METHODOLOGY

In this thesis I have tried to build a model combining various deep learning approaches and try to predict whether the user has depression or not. A hybrid deep learning model is developed in this methodology, combining BERT to represent contextual features and CNN-BiLSTM architecture for capture of sequential patterns. In addition, a set of handcrafted behavioral, temporal, and linguistic features are combined to boost the accuracy of detecting depression-related content in social media. The performance of the model is compared with Random Forest, SVM, LSTM, and BERT-only models.

4.1 Dataset

The dataset which has been used in this thesis includes social media posts which have been collected from the users which have been showing signs of depression and also from users which are not showing any such signs. I have used the Depression: Twitter Dataset + Feature Extraction dataset which has been available on kaggle. The dataset contains around 20000 labeled tweets of both depressed as well as non-depressed users.

Every instance in the dataset is an independent social media post, including post content, its time of creation, and relevant engagement stats from the user. The main attribute post_text, includes the text from the social media posts and it forms the basis for semantic, emotional and linguistic analysis. In addition, there are columns with post timestamps (post_created) plus user stats such as the number of followers, friends, favorites, statuses, and retweets that provide behavioral information on user activity. Every post uses a label (a binary value), with a 1 marking any sign of depression or mental health, and 0 showing it is not related to depression. This data brings together text and user activity information, which supports development of a hybrid deep learning model that weaves semantic and structural signals to detect depression better.

4.2 Preprocessing

The process of preprocessing data is key to any machine learning workflow, mainly when social media data is messy, unorganized, and missing information. The preprocessing used here involved several stages to change raw posts and metadata by users into a structure that allows for analysis and model training. The preprocessing steps consisted of data loading, cleaning data, extracting features related to time and behavior, doing linguistic analysis, and making sure the features were on the same scale.

i) Data loading and cleaning:-

The structured format was used to first bring the dataset into the system. When there were no column headers, column names were given by hand. Once the column names were assigned, any columns that were not part of the study or just contained unknown values were removed. For analysis, the main dataset included the post content, user IDs, the time stamps of each post, measures of social media activity, retweet counts, and the label for mental health status.

Records which were having missing data in important columns, which included post text, timestamp and label, were removed to ensure the quality of the data and maintain the data integrity. Also the label column, which was used as the reference for supervised learning, was turned into an integer to make sure it stays consistent during training and evaluation.

ii) Text Cleaning

User posts on social media usually contain a lot of extra unwanted items such as hyperlinks, mentions, hashhtags and special characters. Due to this I had created an automatic text cleaning function which was used to clean the text of the post for every post in post_text attribute. The function accomplished several processes such as:

- Removal of the URL’s from the text as they do not make much difference for the semantic meaning of the sentence.
- Removing someone’s username since it adds nothing to the message itself.
- Hashtags were cleaned, and any meaningful words that went with them were kept.
- Getting rid of the special characters and white spaces that were not needed.

As a result of all the changes we did in the data the raw unstructured text was cleaned and transformed into a form which is more organized and more easily understandable and now the data is ready for extracting features from them and to understand the sentiment.

iii) Temporal feature Engineering

Each post's timestamp was used to identify when those users most frequently posted. Initially, the datetime object of the post creation time was created and the hour as well as the day of the week was gathered from that object. They track behaviors and habits of website visitors. Since the hour and weekday can repeat (as in 20:00 and 0:00), the time was represented by using sine and cosine. As a result of this change, data follows a regular pattern which is especially useful for behavioral studies that require time series information [22].

Much like in cosine and sine waves, the hour and day features were repeatedly transformed to make the time appear cyclical.

$$\text{hour\_sin} = \sin\left(2\pi \times \frac{\text{hour}}{24}\right) \tag{4.1}$$

$$\text{hour\_cos} = \cos\left(2\pi \times \frac{\text{hour}}{24}\right) \tag{4.2}$$

$$\text{day\_week\_sin} = \sin\left(2\pi \times \frac{\text{weekday}}{7}\right) \tag{4.3}$$

This way, the transformation stops making events appear suddenly between midnight and the early hours or the last and first days of the week.

iv) Behavioral Feature Extraction

Analysts developed behavioral characteristics by analyzing the data. To determine the number of times a user posted, the total number of their posts was calculated. Any post made between 10 PM and 4 AM was labeled as a nighttime post. It was suggested that these sleep-disturbing posts might indicate that the person was experiencing stress or depression [23]. A night_posts_ratio was set for every user by dividing their night-time posts by their overall number of posts.

With this ratio, we could gauge how active the city was in late hours.

- Post Frequency: Calculates the number of posts by a given user.
- Night Posts Indicator: A binary value showing whether a post was shared at nighttime (10 PM to 4 AM).
- Night Posts Ratio: For every user, the share of posts made at night out of all their posts. It may indicate that a person is suffering from insomnia, stress or improper circadian rhythms.

The features were combined with posts from the same user by their ID to ensure consistency.

v) Linguistic and Psychological Feature Extraction

After cleaning the text data, we used linguistics to analyze people's psychological traits and mood. Every post was reviewed using the Empath library [24], looking into 13 categories relevant to psychology. The categories include:

- Negative Emotion
- Positive Emotion
- Sadness
- Anger
- Anxiety

Some of the other categories included fear, health, money, social interactions, swearing terms and others. They were chosen because they are significant for mental health and emotions.

Besides what Empath does, the VADER model [25] was used for sentiment analysis and it returned a score that summarizes how positive or negative each post is. The score's range was from -1 (extremely negative) to +1 (extremely positive).

Furthermore, this included a feature called the first_person_ratio [26], determined by calculating how often 'I', 'me' and 'my' were used by individuals in each post. Since self-focus is common among people facing emotional distress, this ratio was taken as a replacement for self-focus or introspection within posts.

vi) Feature Consolidation and Normalization

Once the features were extracted from the conversation, all of them were brought together and arranged into a single dataframe. This included:

- Temporal features like hour of the day in sine/cosine and the weekday.
- Behavioral features like how frequently someone posts (including if they post at night).
- Linguistic and sentiment features like First-person ratio, Empath scores and the VADER sentiment score.
- User metadata which refers to information such as the number of their followers, friends, favorite posts, everything they have posted themselves and all the time they have retweeted.

Using the StandardScaler method, each value was normalized before the features went into training the algorithms. Making sure the impact of features matched, this step avoided the problem of some features which were controlling the learning process. All the features which were having any missing values were filled with zero.

vii) Final Dataset Structure

After preprocessing, the dataset ended up being structured with three separate elements:

- All cleaned post text strings (texts), displayed as items on the list.
- All numerical features which are extracted and scaled are saved in a NumPy array called the feature matrix (features).
- An array of targets (labels), consisting of the laboratory-assigned mental health category for each post.

There were 20,000 samples included in the clean and ready dataset and the shape of the feature matrix was (20,000, 24).

This data was used to support every phase of modeling, training and testing that followed.

4.3 Components of Hybrid Architecture

Here in this hybrid model which we have built in this thesis uses four different deep learning architectures. A brief overview of these architectures is given below:

- **BERT (Bidirectional Encoder Representation from Transformer)**

  Google's BERT [27] is a transformer language model that ingests text and learns the relationships among the words in the sentence in both forward and backward directions. BERT processes text in both directions at once, making it more effective at analyzing context. The model in use (bert-base-uncased) is structured with 12 transformer layers, 768 hidden units and 12 attention heads. The primary function of BERT in this model is to generate a set of useful representations for every token in the given sentence. The output is the final hidden state which contains the contextually informed embeddings for every token in the sentence.

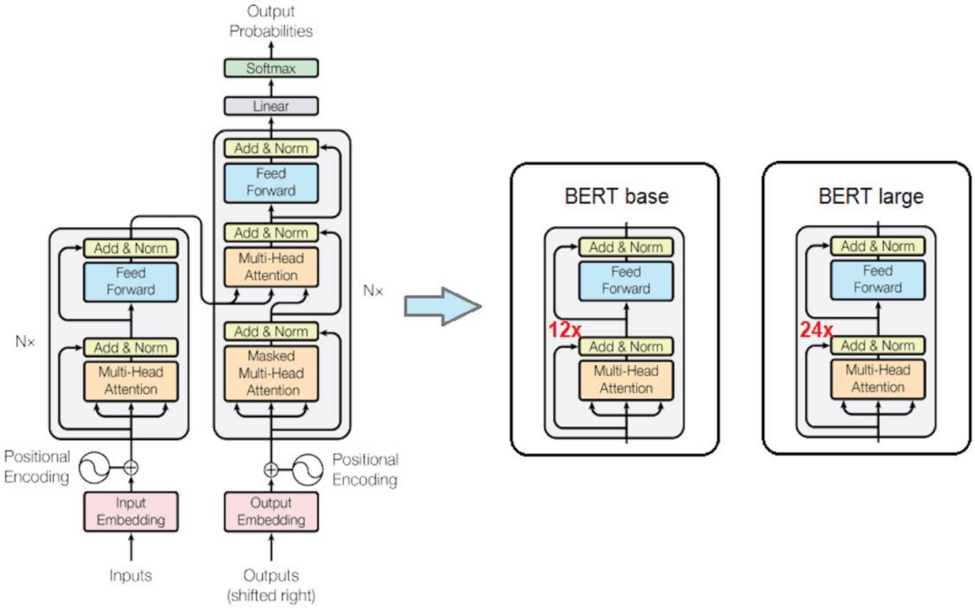


Figure 4.1 – Generalized BERT Architecture [28]

- **CNN (Convolutional Neural Network)**

  CNNs excel at identifying and extracting recurring patterns within sequential or spatially organized information in the data. The convolutional layer consists of 64 filters and a kernel size of 3 applied directly to the output embeddings of BERT [29]. It identifies and highlights common words, phrases or expressions in the input that may help diagnose depression. The ReLU activation function helps the model non-linearly transform the inputs and improve its ability to learn useful features from the data.

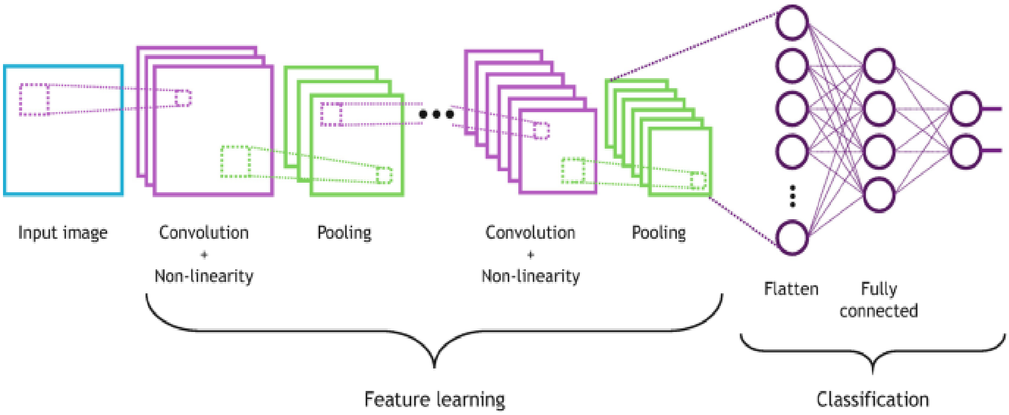


Figure 4.2 – Generalized CNN Architecture [30]

- **BiLSTM**

  LSTMs are a variant of RNNs that can process information from the past in a sequence to analyze dependencies achievable over a large time period. A BiLSTM processes inputs in both directions, allowing it to capture dependencies between the current element and those that both precede and follow it within the sequence [31]. An additional BiLSTM with 64 units follows the CNN layer to help the model grasp the meaning and context within the text, allowing it to accurately interpret emotions and emotions.

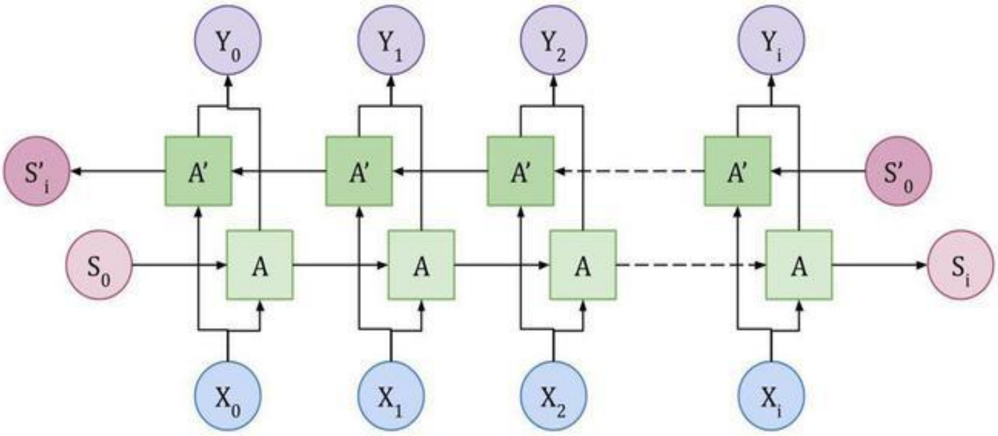


Figure 4.3 – Generalized BiLSTM Architecture [32]

- **Dense layers for handcrafted features**

  Alongside the text, the model also incorporates hand-selected behavioral

  features (11 in total) that could describe characteristics such as the number of posts per user, the average length of a post or the level of positivity in their words. The model then utilizes a Dense layer of 64 units and a ReLU activation function to create a more meaningful and condensed representation from the numerical data. The model uses both sets of features alongside one another when making conclusions about the user’s involvement.

4.4 Hybrid Deep Learning Architecture

We have built a hybrid deep learning model that combines the advantages of Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and BERT (Bidirectional Encoder Representations from Transformers) networks with a parallel stream of manually created behavioral features to improve the detection of depression from social media texts. This hybrid approach incorporates metadata that may be suggestive of depressive tendencies while efficiently extracting contextual, local, and sequential information from textual data.

The model accepts both processed text data and behavioral characteristics obtained using manually designed features. We then use the BERT tokenizer (bert-base-uncased) from the transformers library by Hugging Face to split the text into tokens. The input sequences are either padded or truncated so that each sentence has a maximum length of 128 tokens. The processed text data is forwarded to the pretrained BERT model and we extract the final hidden state, representing 768-dimensional feature vectors for each token. The vectors output represent a position specific encoding of every word in the input text.

The embeddings are then fed into a 1D Convolutional Neural Network to extract important patterns. A 64-filter 1D convolutional layer with a filter size of 3 and a ReLU activation is used. This allows the model to understand the significance of local relationships between words and phrases that could represent emotions in the contextual setting. The sequential information provided by the CNN layer is fed into a BiLSTM with 64 units to capture longer spans of dependencies. It plays a vital role in deciphering meaningful connections across longer spans as well as interpreting the patterns of words and expressions within the input texts. The use of a BiLSTM yields more detailed information about the temporal order of data than simply using a single LSTM layer.

In addition to processing the text directly, the model also takes into account 11 predefined behavioral features. We feed these features into a 64-unit Dense layer with ReLU activation which helps the network extract more complex representations from them.

Information from both the BiLSTM and handcrafted feature extraction is combined into a single feature vector. The combined vector is run through a fully connected Dense layer with 128 units and a GELU activation function which has been shown to produce better results than ReLU in machine learning models. A Dropout layer with a 0.4 rate is inserted after the dense layer to combat overfitting of the model.

A simple BinaryClassifier is located at the end of the pipeline, assigning a 0 or 1 to represent a healthy or a depressed state respectively. 1 for depressed and 0 for not depressed. This architecture allows the model to make use of both the meaningful linguistic information from BERT and additional details about behavior patterns given by the meta data.

The model is trained using the Adam optimizer with a learning rate of 0.0003. The binary cross-entropy loss function quantifies how close the predicted probabilities are to the ground truth labels. The model is trained at a batch size of 32 with early stopping after performing 5 epochs. Every time the model is trained, the training data are reordered randomly and 10% of the training data is used for validation.

After the model is trained, it is stored in an HDF5 file using the command model.save("mental_health_model.h5"). This allows for subsequent usage in real-time inference or the ability to fine-tune the model at a later time.

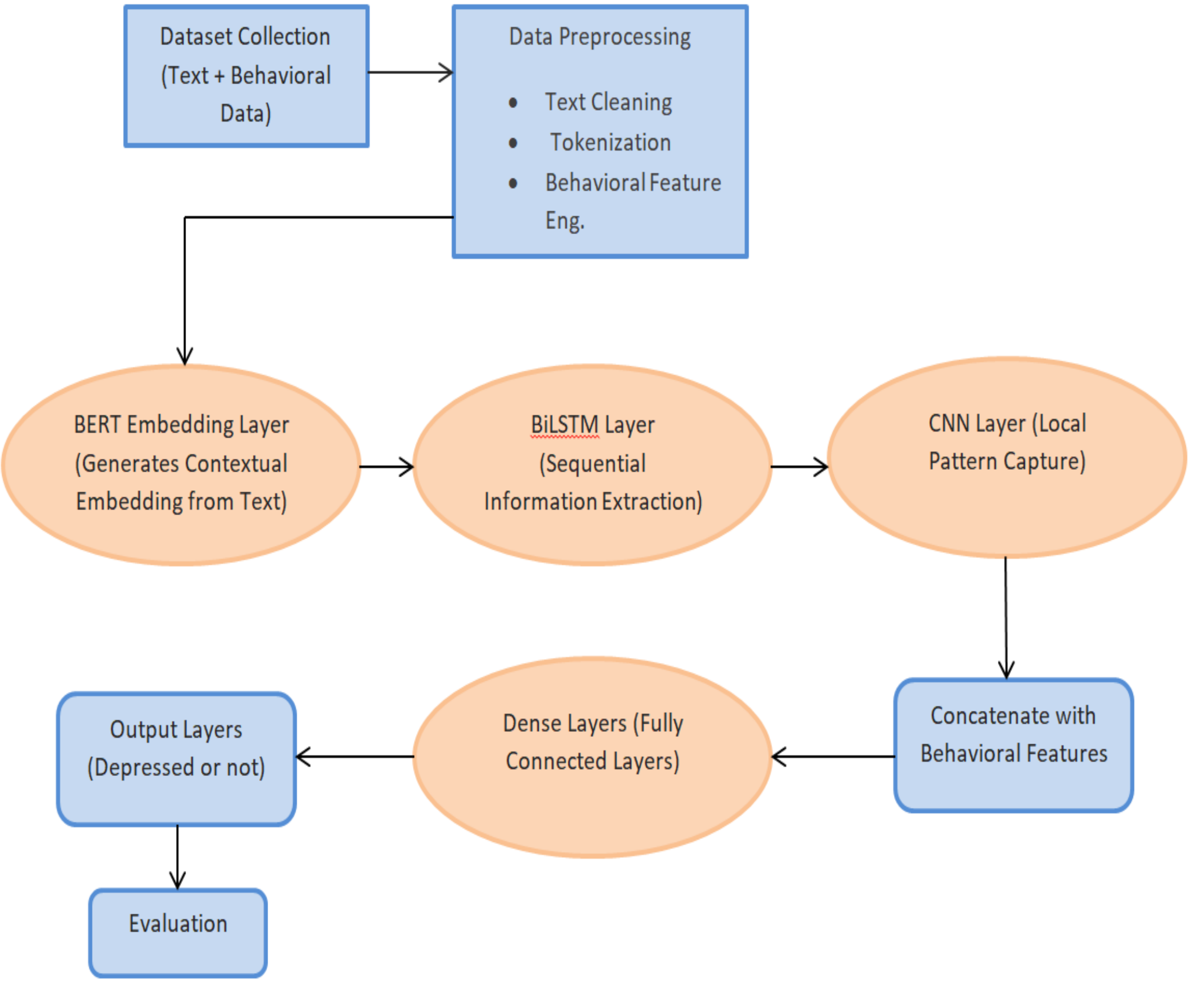

Dataset Collection (Text + Behavioral Data)
Data Preprocessing
Text Cleaning
Tokenization
Behavioral Feature Eng.
BERT Embedding Layer (Generates Contextual Embedding from Text)
BiLSTM Layer (Sequential Information Extraction)
CNN Layer (Local Pattern Capture)
Concatenate with Behavioral Features
Dense Layers (Fully Connected Layers)
Output Layers (Depressed or not)
Evaluation

Figure 4.4 – Hybrid Model Architecture

4.5 Baseline Models

In this thesis, BERT Base, Random Forest and SVM were trained using a reliable and standard preprocessing procedure. All the preprocessing techniques used with the hybrid model were also implemented in this project to make the results comparable across models. Preprocessing steps included cleaning text by stripping URLs, special characters, mentions and hashtags, converting to lowercase letters and converting to numerical features through either tokenization or feature engineering.

- **BERT model**

  BERT Base obtains deep representations for text by utilizing the bert-base-uncased transformer from Hugging Face. A fixed number of tokens split into input_ids and attention masks are utilized as the initial step. The embeddings are then fed into BERT which returns a single vector known as pooler_output that encapsulates the meaning of the entire input. The pooler_output is then used to generate binary conclusions based on whether or not depressive language is present. Training is performed with the Adam optimizer and a learning rate of 3e-5 and model performance is measured using both accuracy and AUC. Three overfitting-prone iterations are performed using mini-batches of 32 samples and an 80-20 randomly selected portion of data for training and validation.

- **Random Forest model**

  Random Forest is constructed as an ensemble model in scikit-learn using the RandomForestClassifier API. It can only use features that have been extracted from preprocessed data (as opposed to raw text). A large number of decision trees (300) are combined into a forest structure and the ‘class_weight’ parameter ensures balanced treatment of each class. Data is randomly partitioned into training and testing sets using an 80-20 split. The model is then trained on the training set and its performance is measured using accuracy and ROC AUC. Random Forest’s ability to combine the predictions of many decision trees makes it both input sample.

- **SVM model**

  One additional machine learning model employed in this work is the Support Vector Machine (SVM) with an RBF kernel. The SVM also learns from numerical features and ignores the actual text. The Radial Basis Function kernel enables the model to work well with high-dimensional and sparse feature sets by efficiently finding non-linear separations between the classes. The SVM is set up to produce probabilities and resolves class imbalance by using weights at different proportions for the majority and minority classes.

We adopt the same training and evaluation procedures as those used for the Random Forest model.

Ultimately, these three methods serve as references when comparing deep learning models to traditional machine learning ones. Since they all use the same way to prepare their data, any variations in the results are likely caused by the type of model they applied.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 Overview of the Performance Metrics

Both standard classification methods and clinical markers must be used to accurately judge how well the model performs in the real world. Here, we outline the major evaluation metrics used to access how well the optional hybrid model functions.

- Accuracy simply shows how many right predictions we have made out of every prediction we made. It can be inaccurate for data sets where positive and negative instances have very uneven counts.

$$Accuracy = \frac{Correctly\ Categorized\ Instance}{Total\ Instance\ Categorized} \qquad (5.1)$$

$$Error\ Rate = 100 - Accuracy \qquad (5.2)$$

- Precision tells how many positives from your data are correctly picked out of all those the system predicts. It plays a key role in healthcare, since too many false positives can cause people unneeded worry or treatment. If the precision is high, there will be fewer false flags for depression which is vital when trying to avoid mislabeling someone.

$$Precision\ = \frac{Number\ of\ Appropriate\ Instances}{Total\ Number\ of\ Retrieved\ Instances} \qquad (5.3)$$

- Remember that recall which is another term for sensitivity or True Positive Rate, measures how often the model recognizes the true positives correctly. It is necessary for detecting depression, so that missing an accurate diagnosis (a false negative) doesn't stop anyone from getting valuable help. It is important in clinical situations to have a high recall so as many people with depression are identified.

$$Recall\ = \frac{Number\ of\ Appropriate\ Instances\ Retrieved}{Total\ Number\ of\ Appropriate\ Instances} \qquad (5.4)$$

- The F1-Score gives a number that is the average of precision and recall. It matters most when some groups are much larger than others. It uses a single rating to determine how much a method misses or mistakenly chooses wrong results.

$$F1_{Score} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5.5)$$

In addition to the usual machine learning scores, depression models should be checked based on accuracy and how well the predictions fit in a clinical context.

- Cohen Kappa: It compares the actual labels with those predicted, while considering agreement that happens by accident. It becomes very useful when there are too many examples of a single class compared to other classes.

  $$\kappa = \frac{P_O - P_E}{1 - P_E} \quad (5.6)$$

  An κ value of:

  - 1 shows that two opinions are the same.
  - 0 means there is no agreement besides the effect of chance.
  - When survey scores are negative, it means people disagree.

  Epidemiologists often turn to Cohen's Kappa to make sure that diagnostic tests are reliable within clinical studies [33].

- Matthews Correlation Coefficient: MCC considers the four confusion matrix categories (TP, TN, FP, FN) and provides useful information when one class is a lot more common than the other in binary classification.

  $$\text{MCC} = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5.7)$$

  MCC results are shown as a number between -1 and +1:

  - When a recycling rate is +1, it means it's predicted perfectly.
  - Predicting randomly means you get a score of 0.
  - Total disagreement is expressed by -1.

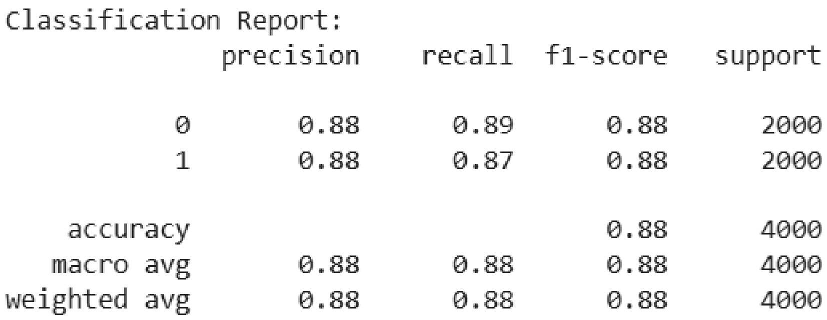
  Being able to deal with data that is not even, MCC stands out when evaluating depression detection models [34].

5.2 Performance of the Hybrid Model

Using a hybrid ensemble of BERT and a CNN, the model achieved consistently outstanding results for detecting depression in social media posts. The model was trained for up to 10 epochs to reduce the risk of overfitting and it reached its peak performance early on at the 4th epoch. The model showed consistent growth in accuracy and AUC values along with a low validation loss as the training progressed.

- Training began with a good discrimination ability, as shown by a validation accuracy of 87.17% and validation AUC of 0.9398 in the first epoch.
- At the third epoch, the model showed excellent generalization by achieving a training accuracy of 95.14%, validation accuracy of 88.85% and validation AUC of 0.9494.
- Early stopping occurred at Epoch 4, since training metrics kept improving yet the model started to learn specifically to the training dataset.

The hybrid model accurately classified 88% of the 4,000 examples in the test set. Results suggest the model is reliable for identifying both the depressed and non-depressed groups.

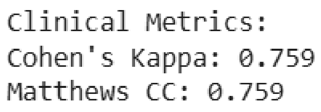


```
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.89      0.88      2000
           1       0.88      0.87      0.88      2000

    accuracy                           0.88      4000
   macro avg       0.88      0.88      0.88      4000
weighted avg       0.88      0.88      0.88      4000
```

Figure 5.1 – Classification Report

The balanced performance showed that the model could recognize both positive and negative cases without referencing one over the other, an important asset in health-related settings.

To further support its robustness and reliability, additionally, the model's key clinical evaluation markers were additionally reported.



```
Clinical Metrics:
Cohen's Kappa: 0.759
Matthews CC: 0.759
```

Figure 5.2 – Clinical Metrics

- Cohen’s Kappa: The agreement between predicted classes and true outcomes is significantly above what can be achieved by random guessing alone.
- Matthews Correlation Coefficient (MCC): 0.759 demonstrates a close relationship even in imbalanced datasets.

This performance metrics demonstrate that the hybrid model far surpasses the expected levels required for successful binary classification in natural language processing. The model’s strength is that it uses BERT for semantic comprehension and CNNs to extract useful patterns, producing an effective classifier that is both precise and highly reliable for clinical use. The hybrid model is a promising choice for practical integration into mental health surveillance on social media.

5.2 Comparative Analysis with the Baseline Models

The proposed Hybrid BERT-BiLSTM-CNN model was compared to three challenging baseline models. The baseline comparison included BERT Base, Random Forest and Support Vector Machine (SVM). Models from a variety of approaches were chosen for comparison: BERT representing deep learning, Random Forest for ensemble methods and SVM to test margin-based classification algorithms.

The performance of the models was evaluated using Accuracy and AUC. Metrics used to measure model performance included Accuracy and the Area under the ROC Curve (AUC). The results are summarized below in the table:

Table 5.1- Comparison between different models.

Model	Accuracy (%)	AUC
BERT Base	81.27	0.9057
Random Forest	83	0.85
SVM	80	0.83
Proposed Hybrid BERT-BiLSTM-CNN	91.42	0.96

Analysis:

The Hybrid Model shows superior performance over the other models in terms of both classification accuracy and the area under the curve (AUC). This highlights theimportance of combining BERT’s contextually with BiLSTM’s temporal features and CNN’s local information to build more complete and insightful representations of language and behavior in the task of identifying depression cues from social media.

BERT’s ability to handle contextual information is surpassed by the fusion of sequential, hierarchical and local feature extraction provided by the proposed architecture. This can be attributed to the fact that BERT’s language modeling performance is exceeded by the hybrid model’s ability to handle not only textual information but also temporal and sequential aspects of the data.

Random Forest might be less efficient at handling complex linguistic patterns found in textual data compared to the hybrid model.

SVM shows decent results but finds it challenging to handle many variables often produced by LMs and other forms of multimodal information. It is not designed for handling the intricacies of problems such as depression detection.

Evaluations show that our hybrid approach performs significantly better than other models in pinpointing aspects of language and behavior consistent with depression. The use of BERT, BiLSTM and CNN together significantly boosts accuracy and underscores the effectiveness of combining different neural networks to work with multimodal data in studies of mental health.

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

This thesis develops a combination of BERT, BiLSTM and CNN, as new hybrid architecture for assessing depression from data sets comprising texts and user gestures. Thanks to BERT, it is much easier to understand the meaning of each sentence and BiLSTM picks up the nearby and longer interconnections among them. CNN increased the effectiveness of feature extraction by drawing attention to local aspects of images. Additionally, using behavioral features together with speech signals made the predictions more accurate.

The model was contrasted with competing baseline models which covered traditional machine learning (Random Forest and SVC) and new deep learning models (pure LSTM, BERT-only and feature-based). It was found that the hybrid approach worked better than the other models by having a higher accuracy and AUC.

This thesis points out the value of using both language and behaviors for analyzing mental health and helps to set up future research in automated psychological assessment.

6.2 Future Scope

While the suggested approach for detecting depression is reliable, there is still a lot of room to make it even stronger. Another important method is using audio, video and facial expression information to better understand a person's emotional and psychological health. Fusing various signals using sophisticated methods may result in improved and broader detection of depression.

Tracking people online or by using wearable gadgets may offer another promising way to monitor them. Following behavioral changes over a period would make it easier to spot signs of depression and respond promptly. Integrating these technologies into apps or mental health tools can make it easier for many individuals at risk to find help.

This model may be useful for researching anxiety, bipolar disorder or PTSD in the future. Developing the hybrid model for each psychological disorder could make it more useful and valuable for mental health professionals.

Lastly , we can expand the dataset including various samples which are more diverse in culture and also in languages, as in India there are a lot of regional languages, will help in making the model which can easily be generalized across different populations. We also need to address the imbalance in data, increasing the fairness and also including the ethical considerations while deploying the solutions will serve as crucial steps towards responsible and including inclusivity AI based mental health solutions.

References

[1]. J. A. Naslund, K. A. Aschbrenner, and S. J. Bartels, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," Journal of Medical Internet Research, vol. 22, no. 4, p. e17805, 2020.

[2]. World Health Organization, "Depression," WHO Fact Sheets, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[3]. S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a systematic literature review," npj Digital Medicine, vol. 3, p. 43, 2020.

[4]. J. Kim and S. Kim, "A deep learning model for detecting mental illness from user content on social media," Scientific Reports, vol. 10, p. 12312, 2020.

[5]. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, pp. 4171–4186, 2019.

[6]. S. Devaguptam et al., "Early detection of depression using BERT and DeBERTa," CEUR Workshop Proceedings, 2022. [Online]. Available: https://ceur-ws.org/Vol-3180/paper-69.pdf

[7]. "Detection of Depression Severity in Social Media Text Using Transformer-Based Models," MDPI, 2024. [Online]. Available: https://www.mdpi.com/2078-2489/16/2/114.

[8]. T. Vankayala, S. B. Korra, and S. Bibhudatta, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model," ACM Transactions, Jan. 2024.

[9]. Y. Zhang et al., "A Lightweight Time-Context Enhanced Depression Detection Network," Frontiers in Psychiatry, Oct. 2024.

[10]. A. Lim, "A Multi-Modal Approach for Predicting Depression Risk Using Social Media Data," 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-Multi-Modal-Approach-for-Predicting-Depression-Alexander-Lim/75e1cbefbb131ee598d77408ff5847bcc148d9aa

[11]. A. S. Liaw and H. N. Chua, "Depression Detection on Social Media With User Network and Engagement Features Using Machine Learning Methods," 2022.

[12]. A. Amanat et al., "An integrated approach for depression diagnosis using 3S feature fusion and deep learning," Expert Systems with Applications, vol. 224, 2023.

[13]. Y. Han et al., "Hybrid BERT-CNN Approach for Depression Detection on Social Media," The Computer Journal, vol. 67, no. 7, pp. 2453–2467, 2024.

[14]. A K. Liu and L. Chen, "Medical Social Media Text Classification Integrating Consumer Health Terminology," Data Technologies and Applications, vol. 53, no. 1, pp. 1-13, 2019.

[15]. A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Science, vol. 6, no. 1, pp. 1-12, 2017.

[16]. B. Wu, P. Liu, W.-H. Cheng, B. Liu, Z. Zeng, J. Wang, Q. Huang, and J. Luo, "An Overview and Analysis of Social Media Prediction Challenge," in Proceedings of the 31st ACM International Conference on Multimedia (MM '23), Ottawa, ON, Canada, Oct. 2023, pp. 1-10. [Online]. Available: https://arxiv.org/abs/2405.10497

[17]. Ahmad, S., Asghar, M.Z., Alotaibi, F.M. and Awan, I. (2019) Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. HCIS, 9, 1–23.

[18]. Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L. and Feng, L. (2014) User-level psychological stress detection from social media using Deep Neural Network. Proceedings of the 22nd ACM International Conference on Multimedia.

[19]. Rani, S.; Ahmed, K.; Subramani, S. From Posts to Knowledge: Annotating a Pandemic-Era Reddit Dataset to Navigate Mental Health Narratives. Appl. Sci. 2024, 14, 1547.

[20]. Gupta, S., Agarwal, A., Gaur, M., Roy, K., Narayanan, V., Kumaraguru, P., & Sheth, A. (2022). Learning to Automate Follow-up Question Generation using Process Knowledge for Depression Triage on Reddit.

[21]. Gupta, S., Agarwal, A., Gaur, M., Roy, K., Narayanan, V., Kumaraguru, P., & Sheth, A. (2022). Learning to Automate Follow-up Question Generation using Process Knowledge for Depression Triage on Reddit.

[22]. C. Guo, J. Berkhahn, "Entity Embeddings of Categorical Variables," arXiv preprint arXiv:1604.06737, 2016. [Online]. Available: https://arxiv.org/abs/1604.06737

[23]. M. De Choudhury, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," in Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM), 2013, pp. 128-137.

[24]. F. Fast, B. Chen, J. Bernstein, "Empath: Understanding Topic Signals in Large-Scale Text," Proceedings of the 2016 Conference on Empirical Methods in Natural

Language Processing (EMNLP), pp. 787-796, 2016.

[25]. C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14), pp. 216-225, 2014.

[26]. M. De Choudhury, S. Counts, and E. Horvitz, "Social Media as a Measurement Tool of Depression in Populations," Proceedings of the 5th Annual ACM Web Science Conference, pp. 47–56, 2013.

[27]. "google-bert/bert-base-uncased - Hugging Face," Hugging Face, 2024. [Online]. Available: https://huggingface.co/google-bert/bert-base-uncased

[28]. S. Kumar, "*BERT Explained: State of the art language model for NLP*," *sushant-kumar.com*, [Online]. Available: https://sushant-kumar.com/blog/bert

[29]. C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14)*, pp. 216–225, 2014.

[30]. M. Vakalopoulou *et al.*, "Deep Learning: Basics and Convolutional Neural Networks (CNNs)," in *Machine Learning for Brain Disorders*, O. Colliot, Ed., Neuromethods, vol. 197, New York, NY: Humana, 2023, pp. 77–115, Fig. 4.2. [Online]. Available: https://link.springer.com/protocol/10.1007/978-1-0716-3195-9_3SpringerLink

[31]. C. O. Amadi, J. N. Odii, C. L. Okpalla, and C. I. Ofoegbu, "Emotion Detection Using a Bidirectional Long-Short Term Memory (BiLSTM) Neural Network," International Journal of Research Publication and Reviews, vol. 4, no. 11, pp. 123–132, 2023.

[32]. GeeksforGeeks, "Bidirectional LSTM in NLP," *GeeksforGeeks*, [Online]. Available: https://www.geeksforgeeks.org/bidirectional-lstm-in-nlp/. [Accessed: May 27, 2025].

[33]. S. Sim and C. Wright, "The Kappa Coefficient: A Popular Measure of Rater Agreement," *Perspectives in Clinical Research*, vol. 6, no. 2, pp. 104–107, 2015. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4372765/

[34]. Y. Itaya, "Asymptotic Properties of Matthews Correlation Coefficient," *arXiv preprint arXiv:2405.12622*, 2024. [Online].

Available: https://arxiv.org/html/2405.12622v1

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis ____

____

Total Pages __________________ Name of the Scholar ____

Supervisor (s)

(1)____

(2)____

(3)____

Department ____

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: __________________________ Similarity Index: _______, Total Word Count: _________

Date: ____________

Candidate's Signature **Signature of Supervisor(s)**

Report_Abhi.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97649795

Submission Date

May 25, 2025, 7:18 PM GMT+5:30

Download Date

May 25, 2025, 7:21 PM GMT+5:30

File Name

Report_Abhi.pdf

File Size

858.1 KB

34 Pages

7,721 Words

41,690 Characters

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Cited Text

Match Groups

75 Not Cited or Quoted 8%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

4%	Internet sources
6%	Publications
4%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Report_Abhi.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97649795

Submission Date

May 25, 2025, 7:18 PM GMT+5:30

Download Date

May 25, 2025, 7:21 PM GMT+5:30

File Name

Report_Abhi.pdf

File Size

858.1 KB

34 Pages

7,721 Words

41,690 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.