Received: October 25, 2020. Revised: December 2, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Bi-LSTM-Based Sentiment Analysis for Identifying Depression Indicators in Twitter Posts

Sharon Susan Jacob[1]*

[1]*School of Data Analytics, Mahatma Gandhi University, Kottayam, Kerala, India*
* Corresponding author's Email: Sharon.s.Jacob@outlook.com

Abstract: Depression remains a significant global mental health issue, and early detection is crucial for timely intervention. Traditional methods often struggle to identify subtle indicators of depression, especially in social media content. This research introduces a novel methodology that combines sentiment analysis with a Bi-directional Long Short-Term Memory (Bi-LSTM) model to identify linguistic markers of depression in social media posts, specifically on Twitter. Sentiment analysis captures emotional cues from text, while the Bi-LSTM model effectively recognizes complex linguistic patterns, enabling a more nuanced detection of depression. The proposed approach overcomes limitations of conventional machine learning models by integrating these techniques, which enhances the identification of both emotional sentiment and linguistic features associated with depressive tendencies. The dataset used for this study is the Kaggle: Depression on Twitter, containing labeled Twitter posts related to depression. The model's performance is evaluated through key metrics, achieving an accuracy of 98.65%, precision of 99.13%, recall of 98.09%, and an F1 score of 98.6%. These results demonstrate the potential of deep learning techniques to improve depression detection in social media content, offering a more reliable and efficient tool for early intervention. This research contributes to the growing field of mental health analysis using social media data, providing a robust framework that can be integrated into mental health support systems to aid in reducing the global burden of depression.

Keywords: Twitter, Sentiment analysis, Depression, Deep learning, bi-LSTM.

1. Introduction

In contemporary society, depression has emerged as a progressively critical public health concern. It is defined as a mental ailment marked by feelings of sorrow and a loss of interest in various aspects of life. At its most severe, depression can escalate to the point of prompting suicide [1]. Notably, research conducted by the World Health Organization (WHO) [2] shows that the global count of individuals experiencing depression surpasses 350 million. Furthermore, the propensity for suicide is markedly higher, over 25 times, among those grappling with depression compared to those without such mental health afflictions [3]. According to recent surveys, when compared with other mental illnesses, depression emerges as the predominant condition in terms of prevalence and impact.

Social networks have ingrained themselves as a significant component of individuals' lives, serving as virtual reflections of users' personal experiences. On these platforms, individuals exhibit a propensity to openly share moments of happiness, contentment, and even moments of sadness [4]. Individuals experiencing depression frequently exhibit an urge to mask their symptoms, either to sidestep disclosing their illness [5] or due to challenges in seeking professional diagnosis [6]. In many instances, these individuals' resort to online social platforms as an outlet to communicate their underlying struggles. This inclination can be attributed, in part, to the willingness to confide in their circle of friends, with the hope that they might extend assistance or guidance. Additionally, some individuals inadvertently drop subtle hints that could allude to their state of depression or the emergence of early depressive signs.

Within the context of modern communication, social networks like Weibo and Twitter have gained immense acceptance as platforms for public discourse on social issues. Twitter, functioning as an open broadcasting medium, enables registered users to engage in discussions and interactions through concise 140-character messages. A recent report from Twitter [7] disclosed the creation of a staggering 1.3 billion accounts, with over 320 million monthly active users and an astounding 140 million users posting more than 450 million tweets daily. People now have a venue to publicly express their feelings and mental states because to the proliferation of these platforms.

The optimal approach to addressing mental illnesses such as depression involves early identification and professional intervention. When diagnosis and treatment are delayed, the condition can often worsen, leading to severe outcomes. This progression can escalate to the point of harboring suicidal thoughts, with individuals concluding that ending their lives is a preferable alternative to seeking help.

Currently, there's no practical method for separating depressed people from those who are not depressed, making it a difficult undertaking. In addition, there aren't enough qualified medical personnel and resources available to treat depression. Because there aren't many efficient approaches for the diagnosis of depression, the majority of the prediction techniques now in use are not accurate. Because of the size of their user bases and the activities they engage in on their individual platforms, social media behemoths like Instagram, Twitter, Facebook, and Snapchat can assist us in anticipating depression. Platforms like Twitter produce 200 billion tweets each year, or 6,000 tweets per second, on average, and they make their data available for open source [8]. This underscores the significance of extending support to individuals exhibiting depressive symptoms and utilizing their social media content as a valuable resource. The resulting insights can serve not only as early indicators for effective intervention but also as supplementary information for mental health professionals.

Recent advancements in Natural Language Processing (NLP), Big Data Analytics, and Artificial Intelligence (AI) offer a potential solution to this challenge. This is particularly relevant in light of the proliferation of various social media platforms, which have proven instrumental in studies focused on mental health detection. Research has highlighted that individuals with conditions subject to societal stigma often turn to the internet for information, interaction, and sharing experiences through social media. NLP is a methodology that has transformed from linguistic theories into computer algorithms with the advancement of computing technology. Sentiment Analysis (SA), a subset of NLP, serves as a technique to scrutinize text or speech and discern the positive and negative attributes embedded within it. Via textual comments and posts on topics of interest, people can express ideas, thoughts, perspectives, and life experiences [9]. By analyzing the negative sentiment ratings in a person's social media posts, this method can be expanded to determine their state of depression.

Understanding "text-related content" has become more difficult as social media content has expanded beyond text to include videos. Elements like emojis, hashtags, and languages beyond English play crucial roles in comprehending an individual's sentiments [10]. Extensive research has demonstrated that appropriately utilizing user-generated content from social networking platforms can enable the early detection of an individual's mental state [11].

Numerous predictive algorithms and optimization methods, including Deep Learning (DL) and Machine Learning (ML), are employed to uncover patterns within data and derive implications from these studies. Within our research, we utilize tweets as input for training and predict depression in users by classifying tweets as either indicative of depression or not. In our study, we utilized a Twitter dataset to undertake the classification of tweets. The proposed methodology centers on tweets, which are concise messages limited to 280 characters. Within this succinct framework, users often express their emotions and thoughts, capturing both personal reflections and observations on global events. Specifically, we labeled tweets as positive for depressed users and negative for non-depressed users, using the Bi-LSTM technique as part of our analysis. This research has the potential to significantly enhance the environmental community's understanding of mental health by utilizing social media data to detect depression, thereby promoting early intervention and support. By fostering a healthier population, it can lead to improved community engagement and resilience, ultimately benefiting both individual well-being and broader societal cohesion in addressing environmental challenges.

The primary contribution of this research lies in its novel integration of sentiment analysis with a Bi-LSTM framework. The sentiment analysis captures emotional cues from text, while the Bi-LSTM model enhances the identification of complex linguistic patterns, enabling a deeper understanding of the interplay between emotional sentiment and

depression-related linguistic features. Unlike traditional methods, which often struggle to detect subtle indicators of depression, the Bi-LSTM model effectively captures these nuances, improving detection accuracy. Moreover, the method addresses the limitations of conventional machine learning models, such as their inability to handle sequential data effectively, by leveraging Bi-LSTM's ability to capture long-range dependencies in textual data. This results in a significant improvement in performance metrics, including accuracy, precision, recall, and F1 score, which surpass many existing approaches in the field. The integration of these advanced techniques allows for more accurate and reliable early detection of depression, which can contribute to more timely interventions and better mental health outcomes. This study's main contribution involves the following:

- A depression detection model has been developed utilizing Twitter data through a combined approach of SA and Bi-LSTM classifiers.
- The model is assessed on the basis of accuracy, precision, recall, and F1-score.

The rest of the paper is organized as follows. In Section 2, a summary of literature is provided, highlighting areas that indicate a need for more investigation. In section three, the methodology is explained in depth. The fourth section goes into great detail about the results that the suggested strategy produced. Finally, a summary of the findings is included in Section five, which gives a conclusion to the paper.

2. Literature review

Over the past two centuries, psychologists have utilized behavioral observation to detect signs of depression in individuals, applying their expertise to recognize subtle cues in emotional and psychological states [12]. With the advent of machine learning (ML), this field has witnessed significant advancements, particularly in healthcare applications [13]. Machine learning has garnered considerable attention for its ability to classify patients into categories of healthy or depressed based on intricate patterns of data. Specifically, researchers have explored the use of statistical ML models to analyze language patterns in texts and conversations, identifying key features that can signal depression. These models classify an individual's writing or speech as depressive or non-depressive, providing a data-driven approach to mental health analysis [14].

The effectiveness of ML methodologies in detecting depression largely depends on the quality and selection of features used for classification. Accurately capturing linguistic and emotional characteristics is essential for reliable prediction. In numerous fields, from healthcare to business and beyond, machine learning has proven to be a versatile tool, empowering researchers to conduct diverse investigations, including those related to mental health. The growing use of ML in depression detection highlights the importance of developing refined models capable of understanding nuanced human behaviors through text analysis. By refining these techniques, researchers aim to enhance the precision of mental health assessments and contribute to better interventions for individuals experiencing depression.

Musleh et al. [15] explored the pervasive influence of depression on various aspects of daily life, such as mood, eating behaviors, and social interactions, with a focus on the lack of mental health awareness in Arabic culture. Despite this, many individuals, including Arabs, express their emotions publicly on platforms like Twitter, often using it as a diary due to its anonymity. While prior research had focused primarily on English-language tweets, their study addressed the gap by developing a model to analyze Arabic tweets for signs of depression. By incorporating a "neutral" label alongside the traditional "depressed" and "non-depressed" categories, the research expanded dataset diversity. The study evaluated several machine learning classifiers and natural language processing (NLP) techniques, with the Random Forest classifier achieving the best performance at 82.39% accuracy. However, the study was limited by its relatively low accuracy, indicating that further refinement of the model and a larger dataset may be needed for improved results. While the inclusion of a "neutral" label helped enhance dataset diversity, the dataset itself may have been insufficiently large or varied to capture the complex linguistic and emotional nuances present in Arabic tweets related to depression.

Kumar et al. [16] underscored the global threat posed by depression, particularly its role as a leading factor in suicide when left undetected, as supported by WHO data. Their study focused on analyzing Twitter posts to identify potential signs of depression, utilizing machine learning (ML) and natural language processing (NLP) techniques. By assigning a sentiment-based numerical score to users' tweets, they achieved 78% accuracy in detecting depression with the XGBoost classifier. Additionally, they combined this sentiment attribute with other linguistic features, such as N-Gram+TF-IDF and Latent Dirichlet Allocation (LDA), which improved the detection accuracy to 89% when using the Support Vector Machine (SVM) classifier. The

research highlighted the critical role of feature selection and the synergistic combination of features in boosting classification performance. While the study demonstrated the importance of feature selection and combining different features for improved classification, a key limitation was the reliance on predefined features. This approach restricted the ability to capture more complex, nuanced indicators of depression in text. The predefined feature set might have missed subtle linguistic patterns and emotional cues that could be crucial for accurate depression detection. The absence of a more dynamic, adaptable feature extraction process limited the depth of the analysis, preventing the study from fully capturing the intricacies of depression as expressed in social media posts. This limitation highlights the need for more advanced techniques that can learn from the data itself, rather than depending solely on predefined features, to better detect the nuanced signs of depression.

Govindasamy et al. [17] addressed the growing concern of depression within the current generation, where many individuals either recognize their condition or remain unaware. The study leveraged the widespread use of social media, particularly Twitter, to detect depressive states using machine learning (ML) algorithms. By analyzing users' social media content, the researchers employed two classifiers, Naïve Bayes (NB) and a hybrid NBTree model, to classify tweets as depressive or non-depressive. The NBTree algorithm achieved an accuracy of 97.31% on a 3000-tweet dataset and 92.34% on a 1000-tweet dataset, while the NB model demonstrated similar performance. The authors primarily focused on the methodology without fully addressing the characteristics of the dataset, such as its size, diversity, and the specific nature of the data (e.g., tweet content, user demographics). This lack of clarity may lead to uncertainties regarding the generalizability of the findings. Additionally, the potential limitations of the data—such as bias in the sample or inconsistencies in labeling depressive content—were not adequately discussed, which could impact the accuracy and applicability of the model when applied to more diverse or larger datasets. The absence of detailed data demonstration also hindered a clear understanding of the model's robustness and its ability to perform well in real-world, varied contexts.

AlSagri et al. [18] investigated the impact of social media platforms such as Facebook, Twitter, and Instagram on mental well-being, focusing on the potential correlation between excessive use and heightened depression rates. The study aimed to utilize machine learning techniques for early detection of depression by analyzing users' network behavior and tweets. A binary classification approach was employed, training classifiers to differentiate between depressed and non-depressed users based on features extracted from their online activities. The research revealed that an increase in the number of features significantly improved accuracy (82.5%) and F-measure scores in identifying depression. The Support Vector Machine (SVM) model showed optimal performance by converting the complex classification problem into a linearly separable one. While the study acknowledged that an increase in the number of features improved the model's performance, it did not demonstrate how these features were selected or how they contributed to the overall model accuracy and F-measure scores. Furthermore, the study highlighted the limitations of the Decision Tree (DT) model, but did not provide a clear example or empirical evidence to substantiate the claim that its comprehensiveness caused challenges in handling new data.

Traditional ML algorithms have demonstrated efficacy in accurately predicting depression based on textual content from social media. In contrast, researchers have explored the potential of DL techniques to extract richer insights from a diverse array of content, including images, videos, unstructured text, and even emojis.

The identification of mental disorders on social media has advanced significantly owing to the use of deep neural networks (DNNs), which include convolutional neural networks (CNNs) and LSTMs [19-20]. DL has achieved outstanding outcomes in NLP tasks including text categorization and SA. Several studies in the literature have focused on using DL models to analyze user content as well as user textual information. Khafaga et al. [21] explored the essential role of social media in emotional expression, highlighting the urgent need for effective depression detection in users' messages. They proposed a novel approach known as Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN), utilizing deep learning to classify Twitter data and identify signs of depression. Their methodology included several preprocessing steps, such as punctuation removal, stop word elimination, lemmatization, tokenization, and stemming. Feature selection was enhanced through Adaptive Particle and Grey Wolf optimization methods. In comparison to established techniques like CNN, SVM, and Minimum Description Length (MDL), the MDH-PWO model achieved an impressive accuracy of 96.86%. Furthermore, the approach did not explore the use of multimodal features, such as emojis,

hashtags, or other contextual elements, which could provide a more holistic understanding of emotional expression in online communication. Consequently, while the model's results were promising, the limitations in feature selection and the narrow scope of emotional cues may reduce the robustness and applicability of the model to a broader range of users.

Amanat et al. [22] addressed the global prevalence and serious consequences of depression, highlighting the urgent need for timely recognition of emotional responses, particularly in the context of widespread social media usage. They proposed a robust model that combined a Long Short-Term Memory (LSTM) architecture, featuring two hidden layers and significant bias, with a Recurrent Neural Network (RNN) that included two dense layers. This innovative approach aimed to predict depression from textual content, providing a means to protect individuals from mental health disorders and suicidal tendencies. By training the RNN with textual data to distinguish markers of depression from semantics and textual phrases, they achieved an impressive accuracy rate of 97.0%, surpassing traditional frequency-based deep learning models and effectively reducing the false positive rate. Specifically, there was no clear explanation of the data collection process, the diversity of the data, or the potential biases that may have influenced the results. Without these details, it is challenging to assess the generalizability of the model across different contexts or populations.

Tejaswini et al. [23] examined the significant impact of depression on individuals and the increasing global prevalence of emotional distress. They utilized natural language processing (NLP) techniques to analyze social media text, aiming to improve depression detection methods. The researchers identified challenges in model representation that impeded accurate text-based detection and introduced the innovative "FasttextCNN with Long Short-Term Memory (FCL)," a hybrid deep learning architecture that leverages the strengths of NLP. The FCL model utilized fasttext embeddings for comprehensive text representation, a convolutional neural network (CNN) for global information extraction, and LSTM for local feature extraction. Testing on real-world datasets revealed that FCL outperformed state-of-the-art methods, achieving a 2.4% and 1% increase in accuracy compared to CNN using word2vec and GloVe embeddings, respectively. Social media users often employ slang, emojis, abbreviations, and context-dependent language, which may alter the interpretation of text, especially in the case of mental health indicators. The lack of consideration for these contextual factors in the FCL model could limit its ability to fully capture the subtleties of depression-related discourse, thereby affecting its overall performance and generalizability. This limitation underscores the need for models to adapt to the dynamic nature of language on social platforms to improve the accuracy of depression detection.

The study by Sri et al. [24] examined the increased utilization of social media platforms like Twitter in Malaysia during the Covid-19 pandemic, specifically investigating the relationship between social media engagement and mental health, with a focus on depression. The researchers employed sentiment analysis (SA) to develop an algorithm that utilized DL and NLP to predict text-based symptoms of depression in urban Malaysian populations during the early stages of the pandemic. The algorithm effectively identified depressive tweets, aiding individuals, caregivers, and healthcare professionals in recognizing signs of mental health decline. The study achieved promising results, reporting an accuracy of 94%, a recall of 0.96, a precision of 0.94, and an F1 score of 0.95. As the study relied on sentiment analysis to detect depressive symptoms in text, the linguistic cues and expressions in rural communities, potentially shaped by cultural, regional, and social factors, were not adequately captured. This could lead to a bias in the algorithm's ability to detect depression across diverse populations.

Despite the notable advancements in utilizing ML and DL techniques for depression detection from social media data, several research gaps persist. Traditional ML approaches exhibit limitations in effectively capturing the temporal and contextual dependencies of textual data, often leading to reduced accuracy in detecting nuanced emotional and linguistic markers of depression. While DL techniques have shown improvements in text representation and feature extraction, many existing models still rely heavily on predefined feature sets, which may overlook more complex, latent patterns within unstructured data. Furthermore, while the integration of sentiment analysis with DL models has demonstrated potential, few studies have fully explored the interaction between linguistic features and emotional cues over extended temporal sequences, which is critical for accurately detecting early indicators of depression. This gap underscores the need for more sophisticated models capable of capturing deeper semantic and emotional nuances through advanced architectures like Bi-LSTM and attention mechanisms, as well as the exploration of cross-linguistic datasets to enhance the generalizability of these models across different cultural contexts.

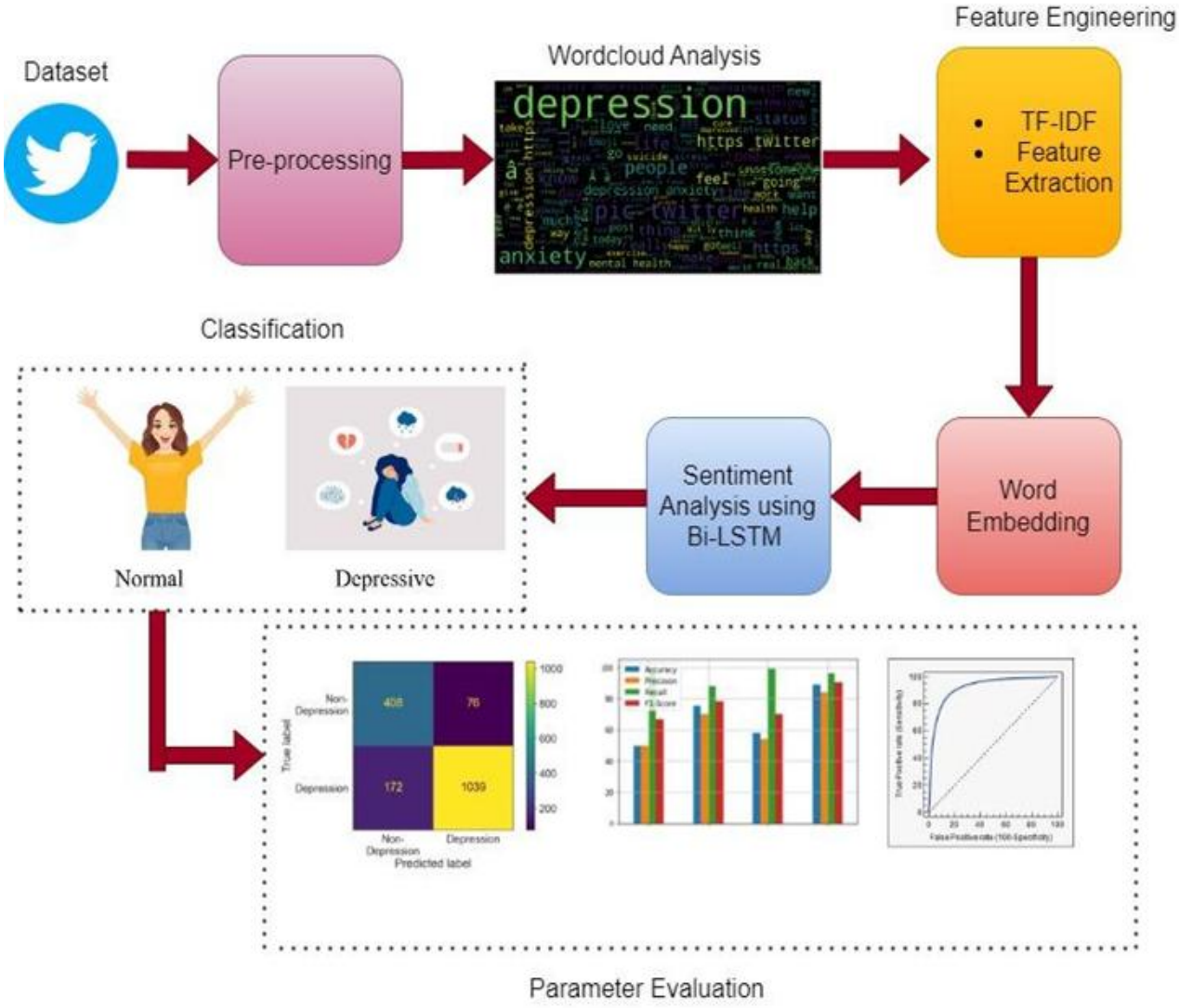

Figure.1 Block diagram of the proposed system

3. Proposed methodology

The proposed method, which combines sentiment analysis with a Bi-directional Long Short-Term Memory (Bi-LSTM) framework, is particularly well-suited for addressing the challenges of detecting depression through social media data, especially Twitter. Sentiment analysis is ideal for capturing subtle emotional cues that are often missed by traditional methods, allowing for a more accurate interpretation of depressive tendencies in text. The Bi-LSTM model further enhances this approach by effectively identifying intricate linguistic patterns over time, addressing the temporal dependencies and context found in user posts. This overcomes the limitations of traditional machine learning techniques, which struggle with understanding the nuanced and sequential nature of language. Existing methods often lack the sophistication to capture the depth and complexity of mental health indicators from text, leading to lower detection accuracy. In contrast, the integration of sentiment analysis with Bi-LSTM allows for a more robust and contextually aware analysis, leading to the superior performance metrics observed in the study. Traditional models also tend to overlook the subtle emotional shifts in text, which is a key limitation in identifying early signs of depression.

The proposed methodology entails a comprehensive approach to detect depression from Twitter data, leveraging DL techniques. The study begins by preprocessing the dataset, including removing duplicates, handling missing values, and cleaning text through special character removal, stemming and lemmatization. Feature extraction encompasses textual features like word frequency, TF-IDF, providing valuable details into the dataset. The data is then split into training and testing sets to facilitate model evaluation. DL models, particularly bi-LSTM is chosen to capture complex patterns in the

text data. While the testing set is used to assess the models, the training set is used to train them, utilizing accuracy, precision, recall, F1-score, and potentially ROC curves for performance assessment. Visualization tools, such as confusion matrices and accuracy-loss plots, aid in interpreting model outcomes. The methodology emphasizes the significance of text-based features in identifying linguistic clues associated with depression, enabling early detection and intervention. Fig. 1 depicts the workflow of the framework.

The proposed algorithm is particularly well-suited for sequential and unstructured text data, such as social media posts, where capturing temporal dependencies and emotional cues is critical. Its ability to process complex linguistic patterns makes it highly effective for detecting depression in diverse, real-time datasets like Twitter, where traditional methods may struggle.

3.1 Dataset description

Evaluating the dataset is of paramount importance in the process of testing and estimating the efficacy of the detection framework. A standardized dataset stands as a fundamental requirement to yield results that are not only accurate but also practically useful. The dataset for the proposed research is sourced from the publicly available online repository on Kaggle, specifically designed to facilitate the analysis of depression indicators on Twitter. This resource can be accessed at Kaggle: Depression on Twitter, providing a comprehensive collection of tweets that serve as a valuable foundation for our investigation into linguistic markers of depression. This dataset encompasses a mixture of tweets, comprising both those indicatives of depression and those that are not as shown in Fig. 2. Within this dataset, tweets expressing depressive sentiments are designated with the label '0,' while tweets without depressive indications are marked with the label '1.' This labeling scheme allows the model to discern between the two categories effectively. The dataset encompasses a balanced distribution of both depressive and non-depressive tweets, facilitating comprehensive analysis. Fig. 3 displays a few representative tweets from the data set along with the labels that were given to them.

3.2 Data pre-processing

The data that was obtained from the sources contains a number of extraneous characters as well as other elements that interfere with the model's architecture. When extraneous data is eliminated from the dataset, model creation becomes simpler. The removal of unnecessary punctuation and stop words is part of the data preparation process for each collected set of data using the Python NLTK package. Emojis and emoticons can provide some crucial information about the sentiment, thus they do not need to be removed for the SA of the text for depression identification. Fig. 4 displays the various text pre-processing procedures.

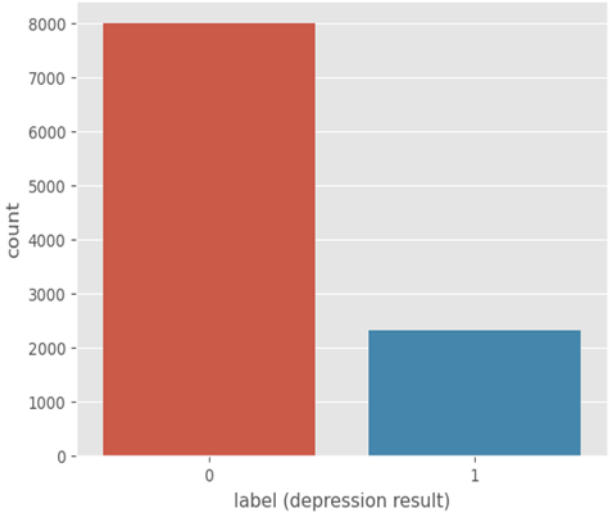

Figure.2 Count of Dataset

	Index	message to examine	label (depression result)
0	106	just had a real good moment. i miss...	0
1	217	is reading manga http://plurk.com/p/mzp1e	0
2	220	@comeagainjen http://twitpic.com/2y2lx - http:...	0
3	288	@lapcat Need to send 'em to my accountant tomo...	0
4	540	ADD ME ON MYSPACE!!! myspace.com/LookThunder	0
...	...	...	...
10309	802309	No Depression by G Herbo is my mood from now o...	1
10310	802310	What do you do when depression succumbs the br...	1
10311	802311	Ketamine Nasal Spray Shows Promise Against Dep...	1
10312	802312	dont mistake a bad day with depression! everyo...	1
10313	802313	0	1

Figure.3 Sample tweets with class labels

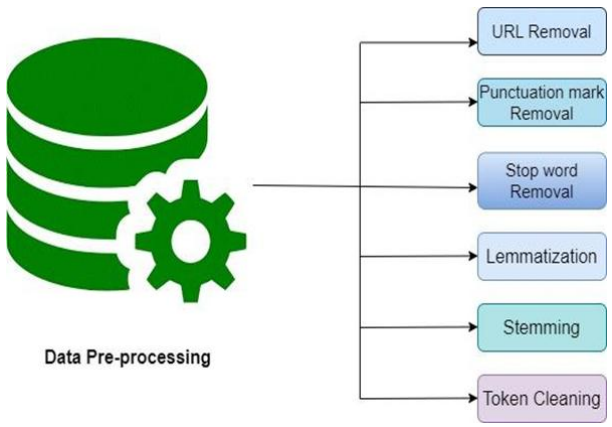

Figure.4 Steps involved in Data Pre-processing

3.2.1. Handling missing values

Before any further processing, it's important to address missing or null values in the dataset. This involves identifying columns or entries with missing data and deciding ways to approach them. Rows with missing values can be removed, or values can be imputed depending on certain standards. It involves identifying and managing instances within a dataset where data points are absent or undefined. Missing values can occur for many reasons, such as mistakes in data gathering, data entry issues, or simply the absence of information for certain observations. In Fig. 5, it's evident that a particular row bearing the index '10313' has been singled out as potentially problematic. Subsequently, this row is removed from the dataset using the '. drop ()' method, with the `inplace=True` parameter set. This configuration ensures that the alteration is directly applied to the original 'data' data frame. The purpose behind this action likely involves the exclusion of the row with index '10313,' which might have been flagged due to concerns related to data quality or other pertinent factors.

3.2.2. Data cleaning

Data cleaning focuses on rectifying inconsistencies, errors, and irrelevant content in the dataset. For Twitter data, this might involve removing special characters, URLs, stopwards, punctuations and irrelevant symbols that don't contribute to the analysis as shown in Fig. 6[25]. Stopwords that don't add to the content of the sentence include "is," "was," "at," "if," and similar words. We utilize the NLTK package, which has a list of stopwords, to eliminate stopwords from the given text.

3.2.3. Lemmatization

Lemmatization is a linguistic and natural language processing technique used to transform words into their base or root forms, known as "lemmas." The goal of lemmatization is to unify words that share a common root, ensuring that different grammatical forms of a word are transformed into the same base form. For instance, the word "running" would be lemmatized to "run," "better" would become "good". To achieve this, lemmatization employs linguistic analysis and dictionaries to identify the lemma associated with each word. This technique considers factors such as tense, plurality, and part of speech to generate the most appropriate lemma for a given word.

	Index	message to examine	label (depression result)
0	106	just had a real good moment. i mississsssss hi...	0
1	217	is reading manga http://plurk.com/p/mzp1e	0
2	220	@comeagainjen http://twitpic.com/2y2lx - http:...	0
3	288	@lapcat Need to send 'em to my accountant tomo...	0
4	540	ADD ME ON MYSPACE!!! myspace.com/LookThunder	0
...	...	...	...
10308	802308	Many sufferers of depression aren't sad; they ...	1
10309	802309	No Depression by G Herbo is my mood from now o...	1
10310	802310	What do you do when depression succumbs the br...	1
10311	802311	Ketamine Nasal Spray Shows Promise Against Dep...	1
10312	802312	dont mistake a bad day with depression! everyo...	1

Figure.5 Data after missing value handling

	Index	message to examine	label (depression result)	urlsRemoved	punkt	stopWord_Removed
0	106	just had a real good moment. i mississsssss hi...	0	just had a real good moment. i mississsssss hi...	just had a real good moment i mississsssss hi...	real good moment mississsssss much
1	217	is reading manga http://plurk.com/p/mzp1e	0	is reading manga	is reading manga	reading manga
2	220	@comeagainjen http://twitpic.com/2y2lx - http:...	0	@comeagainjen -	comeagainjen	comeagainjen
3	288	@lapcat Need to send 'em to my accountant tomo...	0	@lapcat Need to send 'em to my accountant tomo...	lapcat Need to send em to my accountant tomo...	lapcat Need send em accountant tomorrow Oddly ...
4	540	ADD ME ON MYSPACE!!! myspace.com/LookThunder	0	ADD ME ON MYSPACE!!!	ADD ME ON MYSPACE	ADD ME ON MYSPACE
...	...	...	...	...	...	...
10308	802308	Many sufferers of depression aren't sad; they ...	1	Many sufferers of depression aren't sad; they ...	Many sufferers of depression aren t sad they ...	Many sufferers depression sad feel nothing per...
10309	802309	No Depression by G Herbo is my mood from now o...	1	No Depression by G Herbo is my mood from now o...	No Depression by G Herbo is my mood from now o...	No Depression G Herbo mood done stressing peop...
10310	802310	What do you do when depression succumbs the br...	1	What do you do when depression succumbs the br...	What do you do when depression succumbs the br...	What depression succumbs brain makes feel like...
10311	802311	Ketamine Nasal Spray Shows Promise Against Dep...	1	Ketamine Nasal Spray Shows Promise Against Dep...	Ketamine Nasal Spray Shows Promise Against Dep...	Ketamine Nasal Spray Shows Promise Against Dep...
10312	802312	dont mistake a bad day with depression! everyo...	1	dont mistake a bad day with depression! everyo...	dont mistake a bad day with depression everyo...	dont mistake bad day depression everyone em

Figure.6 Data after data cleaning

	Index	message to examine	label (depression result)	lemmatizedRows	stemmedRows	cleanTokens
0	106	just had a real good moment. i mississsssss hi...	0	real good moment miss much	real good moment miss much	real good moment miss much
1	217	is reading manga http://plurk.com/p/mzp1e	0	reading manga	read manga	read manga
2	220	@comeagainjen http://twitpic.com/2y2lx - http:...	0	comeagainjen	comeagainjen	comeagainjen
3	288	@lapcat Need to send 'em to my accountant tomo...	0	lapcat Need send em accountant tomorrow Oddly ...	lapcat need send em account tomorrow oddli i e...	lapcat need send account tomorrow oddli even r...
4	540	ADD ME ON MYSPACE!!! myspace.com/LookThunder	0	ADD ME ON MYSPACE	add ME ON myspac	add myspac
...	...	...	...	...	...	...
10308	802308	Many sufferers of depression aren't sad; they ...	1	Many sufferer depression sad feel nothing pers...	mani suffer depress sad feel noth persist nag ...	mani suffer depress sad feel noth persist nag ...
10309	802309	No Depression by G Herbo is my mood from now o...	1	No Depression G Herbo mood done stressing peop...	No depress G herbo mood done stress peopl deserv	depress herbo mood done stress peopl deserv
10310	802310	What do you do when depression succumbs the br...	1	What depression succumbs brain make feel like ...	what depress succumb brain make feel like neve...	what depress succumb brain make feel like neve...
10311	802311	Ketamine Nasal Spray Shows Promise Against Dep...	1	Ketamine Nasal Spray Shows Promise Against Dep...	ketamin nasal spray show promis against depres...	ketamin nasal spray show promis against depres...
10312	802312	dont mistake a bad day with depression! everyo...	1	dont mistake bad day depression everyone em	dont mistak bad day depress everyon em	dont mistak bad day depress everyon

Figure.7 Data after Lemmatization, Stemming and Token cleaning

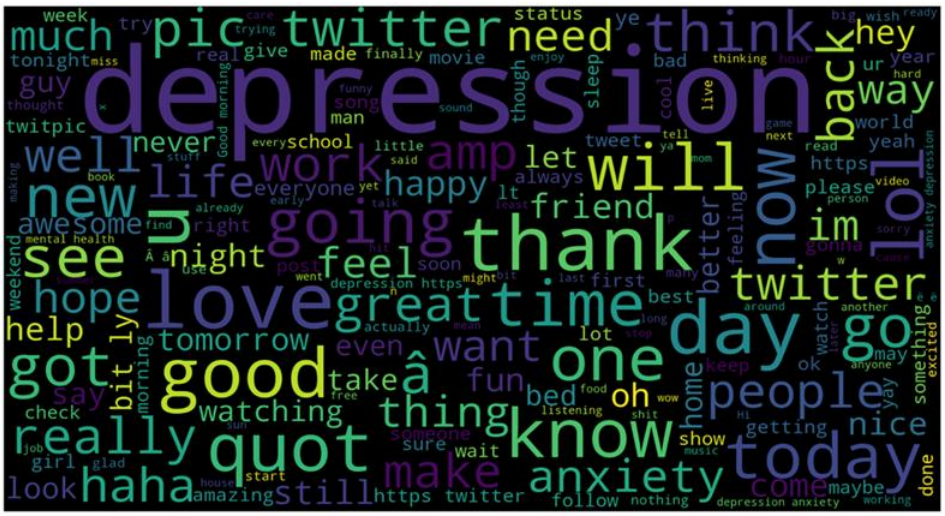

Figure.8 Word cloud for random tweets

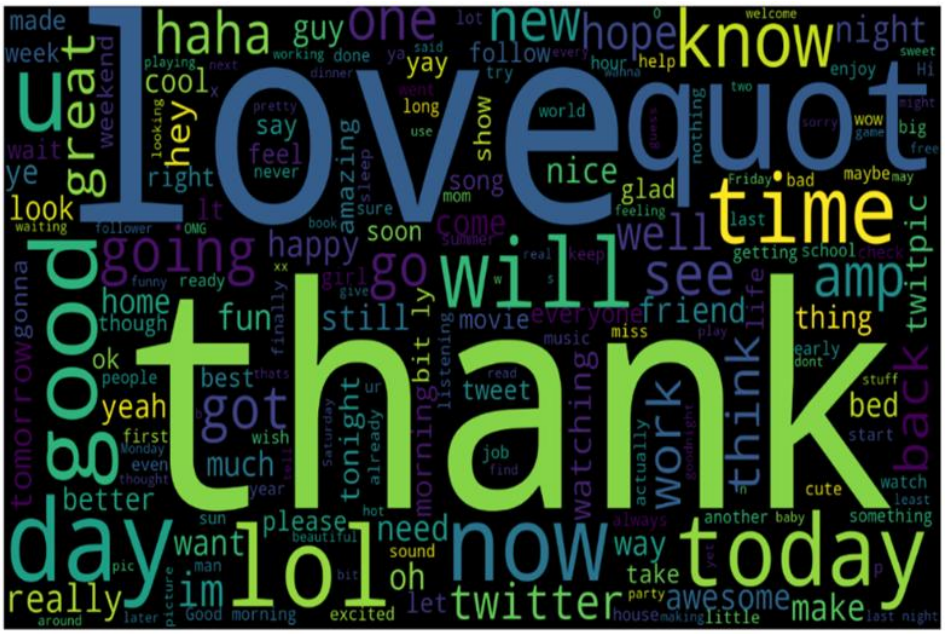

Figure.9 Word cloud for non-depressive tweets

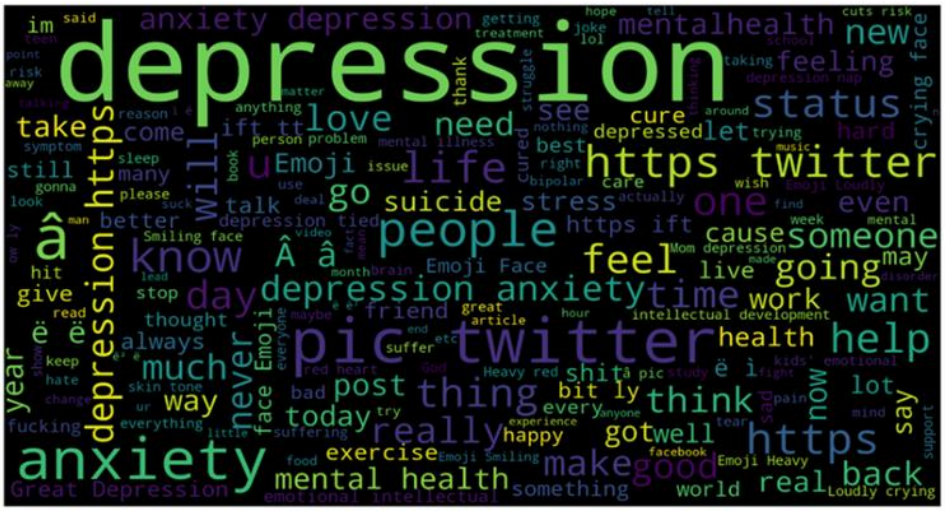

Figure.10 Word cloud for depressive tweets

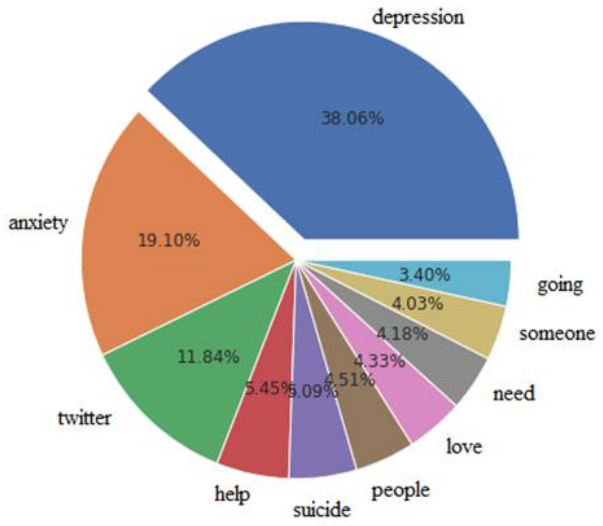

Figure.11 Most Common word count in depressive tweets

3.2.4. Stemming

Stemming is a text processing technique commonly employed in NLP to simplify words down to their root form, known as the "stem." In this process, affixes like prefixes and suffixes (−ize, −ed, −s, −de, etc.) are systematically removed from words while retaining the core morphological unit by Porter Stemmer. The primary objective of stemming is to consolidate words that share the same root, which may have been inflected or conjugated differently due to tense, pluralization, or other grammatical reasons. For instance, after undergoing stemming, words like "running," "runner," and "runs" would all be reduced to their common root "run."

3.2.5. Token cleaning

A token is a fundamental unit of text that has been extracted from a larger body of text [26]. After stemming, there might still be remnants of characters that are not relevant to the analysis, or there could be situations where stemming produces incomplete or incorrect words due to the aggressive nature of the process. Hence, the clean token after stemming step involves a second round of cleaning to address these potential issues.

The cleaning of stemmed tokens involves steps such as removing any residual punctuation marks, special characters, and non-alphabetic characters that might have been introduced during stemming as shown in Fig. 7. Additionally, any very short or very common words that might have survived the stemming process but don't contribute substantially to the analysis are often filtered out. For instance, after stemming, a token like "happi" might result from the word "happiness." However, by performing clean token after stemming, this token could be further simplified to "happy," eliminating the residual "i" and enhancing the token's clarity and alignment with its base meaning.

In order to visualize the data, a word cloud analysis was performed on the pre-processed tweets in the datasets. Word cloud highlights the frequency of words within a dataset. It presents words in varying sizes based on their occurrence, where larger words indicate higher frequency. Fig. 8 illustrates a collection of random tweets sourced from the dataset.

Fig. 9, highlights the occurrence of words within the dataset that originates from random tweets devoid of any discernible traits indicating depression. On the other hand, Fig. 10 distinctly presents the existence of specific words that convey negative emotions. The presence of these words within an individual's language serves as an indicative marker of a propensity towards depressive inclinations.

Analyzing the word count distribution plays a pivotal role in identifying key terms that potentially hint at depressive tendencies or the emotional state of individuals. Disparities in word counts between the two categories—depressive and non-depressive content—underscore linguistic nuances that hold significance for constructing robust classification models. The prevalence of certain words carries substantial value as indicators of emotions, viewpoints, and apprehensions conveyed by users on the platform. For instance, words linked to emotional states, seeking assistance, or expressing negativity could manifest diversely in the language of individuals grappling with depression as opposed to those not affected. This distinction forms the foundation for discerning and categorizing text data accurately.

As depicted in Fig. 11, the conspicuous prominence of words such as "depression" and "anxiety" within the word cloud of depressive tweets vividly reflects how individuals candidly articulate their emotional struggles and mental health challenges on platforms like Twitter. Furthermore, the quantified occurrences of terms like "help" and "love" potentially hint at the existence of supportive or positive sentiments within the dataset.

3.3 Feature engineering

The process of choosing, manipulating, or combining pertinent characteristics from raw data to improve the effectiveness of DL models is known as feature engineering. It involves extracting meaningful information from the data to improve the model's ability to understand patterns and make accurate predictions. This can include selecting the most relevant attributes, creating new features, and

transforming existing ones. In the context of text analysis, feature engineering can encompass techniques as follows:

3.3.1. TF-IDF vectorization

TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is a method in NLP that converts text data into numerical vectors. It displays the weighting of each word in a document with relation to the corpus as a whole. Words are given weights based on the frequency with which they occur in the content (term frequency) and the rarity with which they occur throughout the whole data set (inverse document frequency). The value of words in documents is thus quantified by a numerical representation. As a result, this method returns the most possible number of words, which in our framework is 10,313 with an index for every word.

The `fit_transform () ` method fits the vectorizer to the data and transforms the text data into a sparse matrix containing the TF-IDF values. The `get_feature_names_out () ` method retrieves the names of the words in the vectorized data. Finally, the resulting TF-IDF matrix is converted to a dense array and organized into a panda DataFrame, where each cell denotes the TF-IDF score of a specific word in a particular tweet. This matrix forms the basis for further analysis, such as training machine learning models, to predict sentiment or classify tweets based on their content.

3.3.2. Feature extraction

Feature extraction can be considered a form of feature engineering. It involves transforming raw data into a format that is more suitable for analysis, often by selecting or aggregating relevant attributes. It involves generating additional features from the text data, like word count, average word length, or the presence of specific keywords.

3.4 Word embedding

A popular technique for obtaining word correlations and relationships from a sizable text corpus is word embedding. Conversely, within an embedding, words are encoded as dense vectors, wherein each vector encapsulates the word's projection into a continuous vector field [27]. This positioning of a word is determined by the context in which it is employed, influenced by the neighboring words, and is acquired through the text contained within the vector space. Essentially, a word's embedding encapsulates its specific location within the acquired vector space. Real number vectors are used to represent these words. Such a model might aid in the discovery of like words or offer a list of words for a phrase. For generating word embeddings from unprocessed text, different approaches have been developed.

For each pre-processed data point, numerical vectors were generated utilising the "Word embeddings" technique. To generate word indexes, we first turned all of the example text's words into sequences. These indices are retrieved using the Keras text tokenizer. Every term has had its tokenizer's index checked to make sure it doesn't receive a zero value, and the vocabulary limit has been modified as appropriate. After that, every word in the dataset is given a special index, which is then utilized to create numerical vectors of every text sample. The length of each tweet is first measured in order to create text sequences. Fig. 12 shows a histogram of tweet counts as word length increases, and it is clear that most of the training set's tweets have fewer than 25 words.

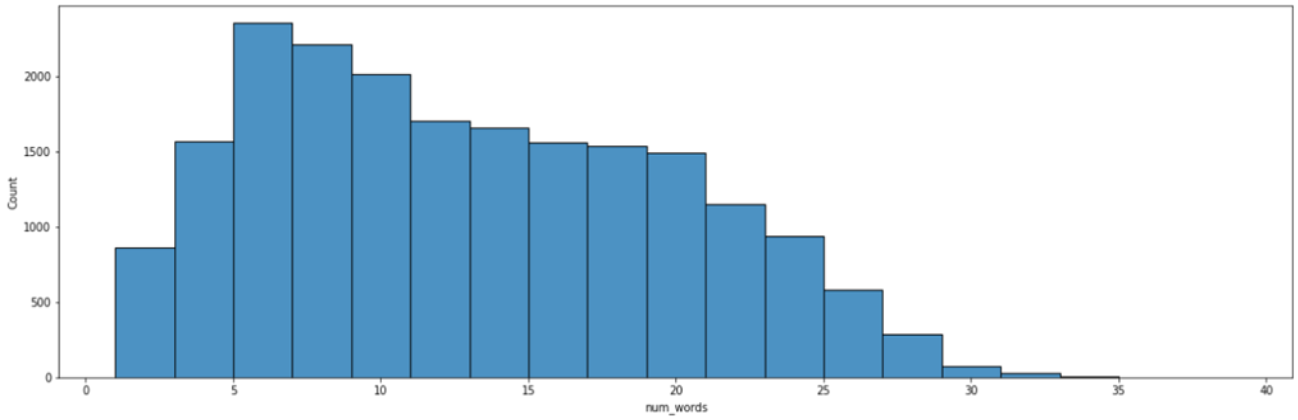

Figure.12 Generation of text sequences

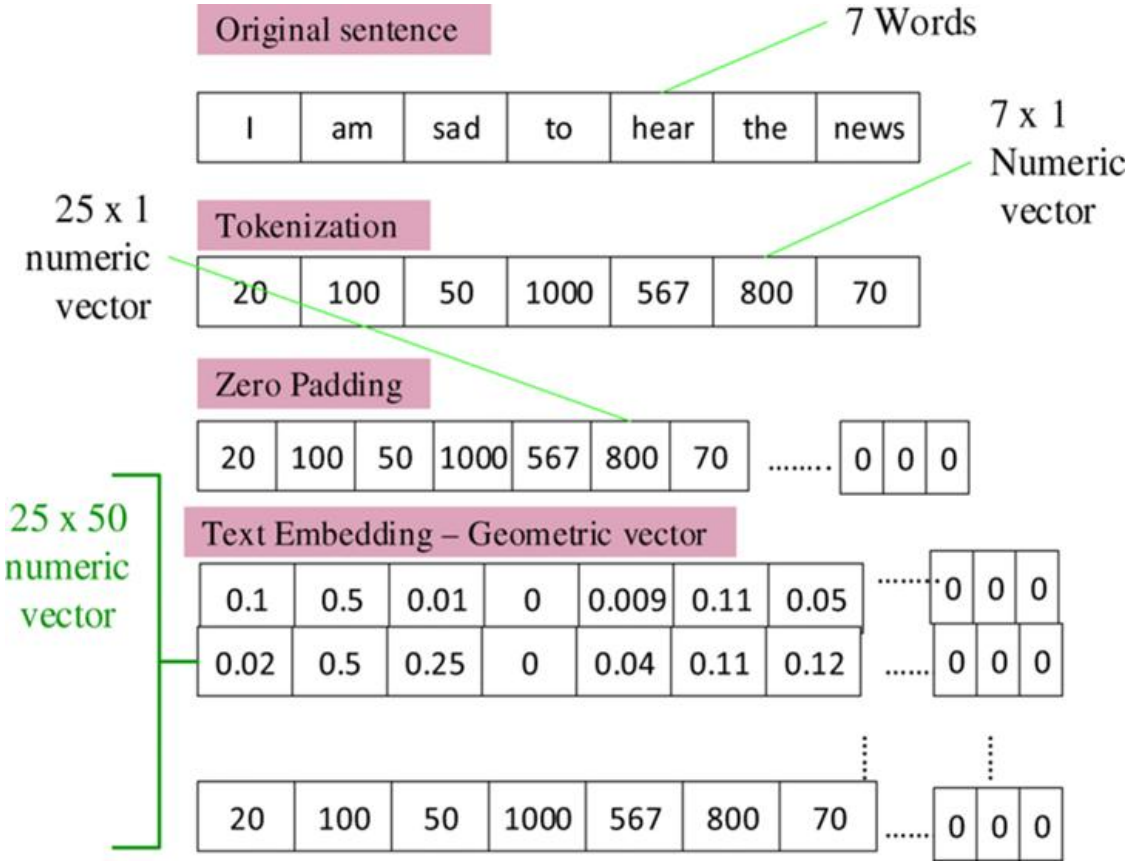

Figure.13 Embeddings of words for a single tweet

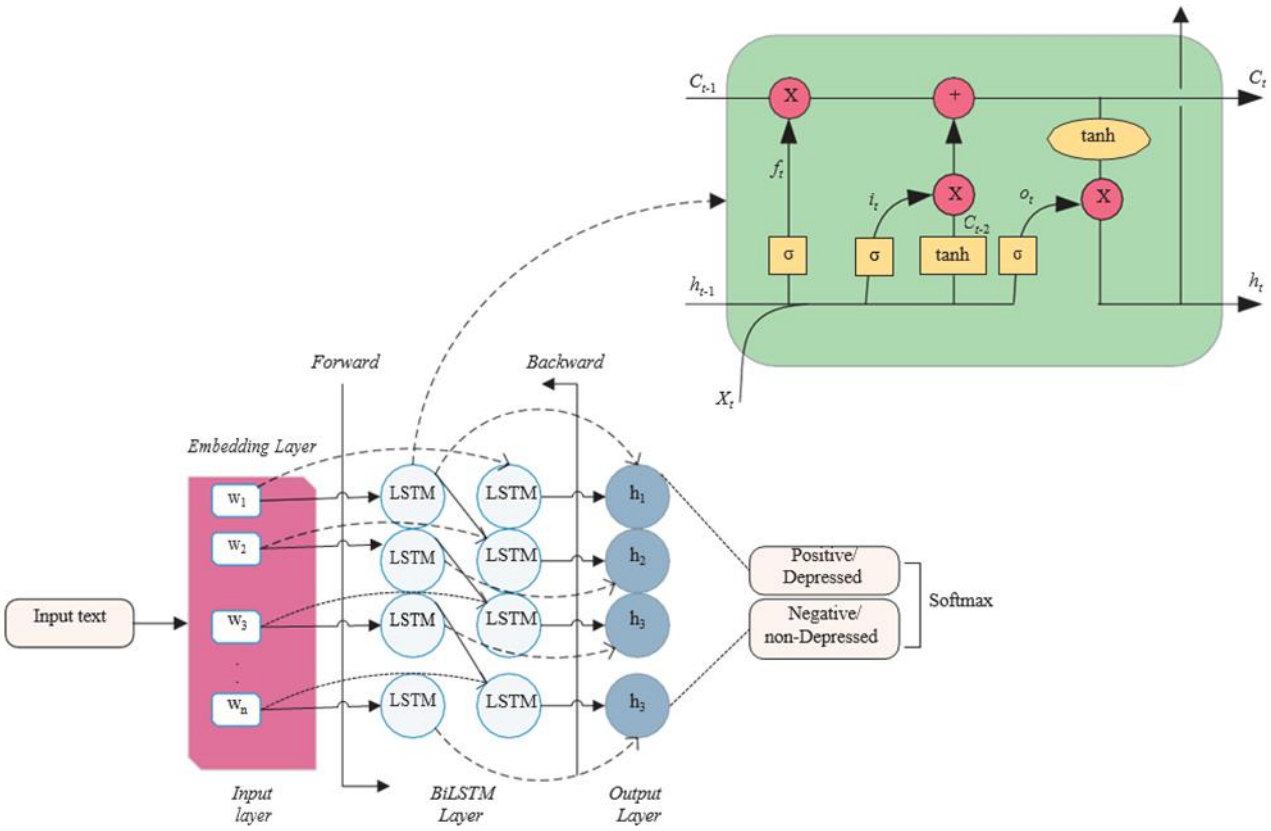

Figure.14 Bi-LSTM Architecture

As a result, zero padding is used and text sequences are transformed into integer sequences. Since most tweets in the dataset are 25 characters or less in length, the maximum sequence length (number of words) has been set at 25. Additionally, 5 terms are eliminated since they appear in very few tweets and contribute zeros to the vector sequence, delaying model training and degrading performance overall. Fig. 13 shows the procedure for creating word embeddings for a tweet. An embedding matrix is created, where the dimensions are determined by the number of unique words (Max_unique_words) and the length of the vector (Embedding_dim), which is set to 300. This matrix is initialized with zeros and serves as the foundation for representing words in a numerical format. The top 10,313 most distinctive words are each transformed into a 300-D vector within the vector space.

For every unique word, its corresponding vector is placed as a row within the embedding matrix. This embedding matrix thus encapsulates the unique words in a 300-dimensional feature space, effectively capturing the nuances of their meanings. Within our bi-LSTM architecture, an embedding layer with a length of 50 is employed. For each tokenized vector, this layer produces an output embedding vector measuring 25 x 50. The purpose of this layered neural network is to learn and reconstruct the linguistic contexts of texts. By feeding in a substantial corpus of text as input (EMBEDDING_FILE), the network creates a vector space with dimensions often ranging in the hundreds. Each word within the dataset is allocated a unique vector within this space, enabling the capture of semantic relationships and contextual understanding. Ultimately, the training and testing classification tasks are executed by employing the bi-LSTM Classifier, culminating in the final stages of the process.

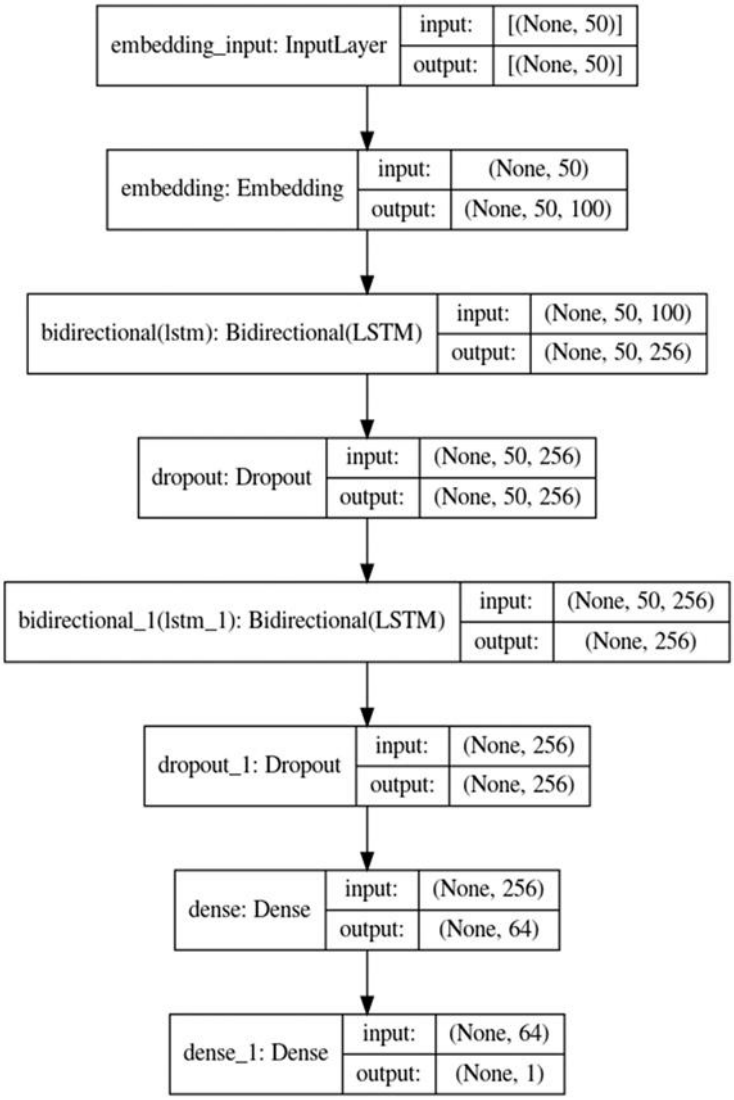

Figure.15 Proposed Model Architecture

3.5 Proposed methodology of sentiment analysis using Bi-LSTM

An NLP approach called sentiment analysis seeks to identify the emotional undertone or sentiment contained in text data. It involves classifying text into groups such as negative, positive, or neutral based on the conveyed emotions. In the context of the study, SA plays a pivotal role in deciphering the emotional content of tweets and identifying signs of depression. In the context of mental health, people often express their emotions, struggles, and experiences through their online interactions, including social media platforms like Twitter. To initiate the process of sentiment analysis, we initially employed the SentiWordNet library within the Python programming framework.

In this study, SA is performed using a specialized deep learning approach known as Bi-LSTM. The Bi-LSTM architecture involves a specialized neural network structure that leverages the power of RNNs to capture sequential information in both forward and backward directions as shown in Fig. 14. This architecture is well-suited for analyzing textual data, such as tweets, where the order of words is crucial for understanding context and sentiment. The Bi-LSTM network consists of two LSTM layers: one that processes the sequence from left to right (forward direction) and the other from right to left (backward direction). The outputs of these two LSTM layers are concatenated to form the final output. This bidirectional processing allows the model to capture contextual information from both the preceding and succeeding words, which is essential for understanding the full context of depressive language.

Each LSTM unit comprises a set of gates: the forget gate (f_t), the input gate (i_t), the cell state (C_t), and the output gate (o_t). These gates are defined mathematically as in Eqs. (1) to (4):

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

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

$$C_t = f_t.C_{t-1} + i_t.tanh(W_c.[h_{t-1}, x_t] + b_C) \quad (3)$$

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

Where, x_t is the input at time stamp t, h_{t-1} is the hidden state from the previous time step, C_{t-1} is the memory cell from the previous time step, W_f, W_i, W_C, W_o are weight matrices and b_f, b_i, b_c, b_o are the bias terms.

The input layer takes tokenized sequences of words (textual data) as input. Each word is represented as an index in the vocabulary.

The tokenized input sequences are fed into an embedding layer, where words are transformed into dense vectors that capture their semantic meanings and contextual relationships expressed as in Eq. (5).

$$Embedding\ (w_i) = E_{w_i} \quad (5)$$

where, E_{wi} represents the embedding vector for the i^{th} word.

The embedded sequences are then passed to the Bi-LSTM layer. The outputs from both directions are concatenated to capture both past and future dependencies of each word. This is vital because the emotional tone of a word can be impacted not only by its preceding words but also by those that follow. Mathematically, the forward and backward LSTM operations are given by:

Forward LSTM:

$$h_t \rightarrow = LSTM_{forward}(\ h_{t-1} \rightarrow\ , E_{wt}) \quad (6)$$

Backward LSTM:

$$h_t \leftarrow\ =\ LSTM_{backward}(\ h_{t+1} \leftarrow, E_{wt}) \quad (7)$$

The output at each time step is given by:

$$h_t = [\ h_t \rightarrow\ ;\ h_t \leftarrow\ ] \quad (8)$$

where [· ; .] denotes concatenation.

The final outputs from the Bi-LSTM layer are fed into a fully connected layer followed by a Softmax activation function to produce class probabilities. This layer assigns a probability score to each class (depressive or non-depressive) based on the learned features from the input sequence expressed as in Eq. (9).

$$Output = SoftMax\ (W.h_t + b) \quad (9)$$

where b is the bias vector and W is the weight matrix.

To adapt Bi-LSTM for depression detection, labeled data containing both depressive and non-depressive text is required. The Bi-LSTM network is trained on this data, learning to recognize patterns associated with different emotional tones. During training, the network adjusts its internal parameters to minimize a loss function that quantifies the difference between predicted sentiments and actual labels. Fig. 15 displays the architecture of the model.

Table 1. Hyperparameter specifications

Hyper Parameters	Values
Batch size	32
Learning rate	0.001
dropout	0.2
Optimizer	Adam
Epochs	50
Activation	SoftMax
Loss function	binary cross entropy

3.6 Hardware and software setup

The proposed model was implemented after the dataset had undergone preparation. This prepared dataset was divided into two separate subsets: an 80% portion designated for training and a 20% portion allocated for testing. The design, training, and evaluation of the model were performed utilizing bi-LSTM architecture within the Google Collaboratory environment. Throughout the entire process, Python programming language, Keras and TensorFlow framework was employed. The specifics of the hyperparameters employed in this study are presented in Table 1, offering insights into the settings that guided the model's training and performance evaluation.

4. Results and Discussion

The accuracy and loss plots are crucial visual representations that provide insights into the performance and learning process of the depression detection model as shown in Fig. 16 (a). The accuracy plot provides a visual illustration of the model's accuracy as it undergoes training iterations on both the training and validation datasets. Accuracy quantifies how closely the model's predictions match the data's actual labels.

Throughout the training process, the accuracy plot offers insights into the model's proficiency in distinguishing between depressive and non-depressive tweets. During early epochs, the ideal scenario is for both the training and validation accuracy to rise concurrently. This trend signifies that the model is successfully acquiring patterns that apply beyond the training data, thus demonstrating its capability to generalize to new, unseen data. This is a critical aspect as it implies that the model is not merely memorizing the training data, but rather comprehending the underlying patterns that distinguish depressive and non-depressive content.

The loss plot provides a graphical representation of the loss, a numerical measure that signifies the discrepancy between the predicted outcomes produced by the framework and the actual true labels of the data as shown in Fig. 16 (b). The loss function plays a pivotal role in guiding the model's optimization process throughout the training phase. Its purpose is to drive the model towards the reduction of errors in its predictions by penalizing deviations from the actual labels. Consequently, as the training progresses, the loss ideally exhibits a downward trajectory, indicating that the model is iteratively improving its predictions by reducing the difference among predicted and actual values.

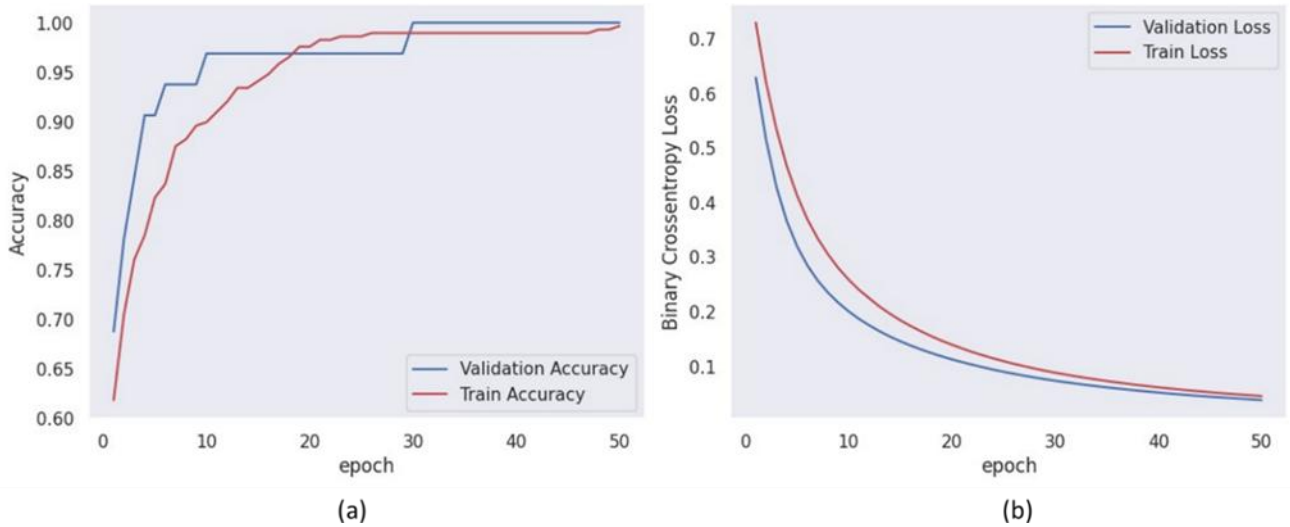

Figure. 16: (a) Accuracy plot and (b) Loss Plot

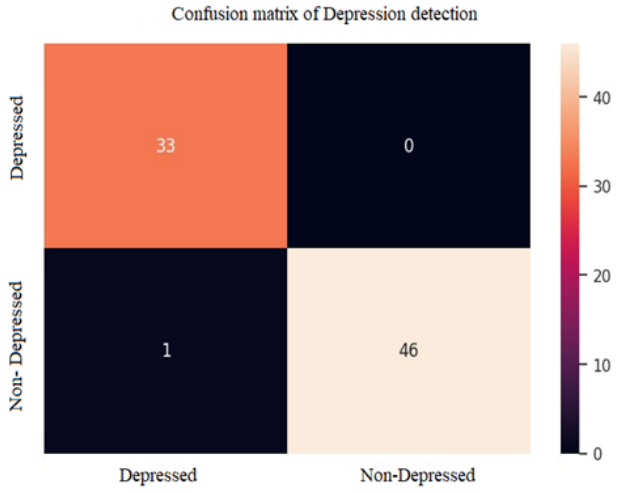

Figure.17 Confusion matrix

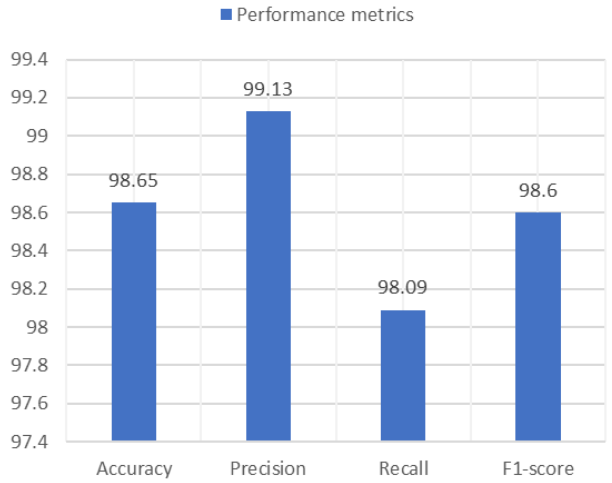

Figure.18 Performance evaluation

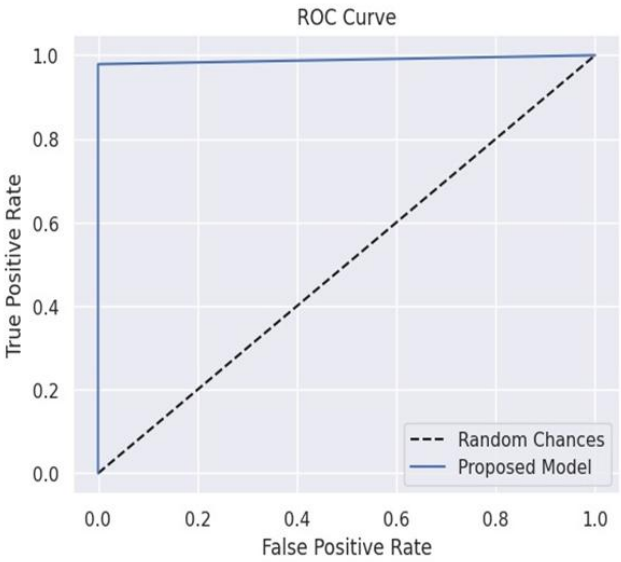

Figure.19 ROC curve

To measure the effectiveness of the framework, a confusion matrix and relevant performance metrics are utilized. A confusion matrix is a visualization that provides a clear overview of the model's classification performance as shown in Fig. 17. It empowers researchers to make informed decisions for enhancing the accuracy and reliability of the depression detection model, contributing to its real-world application in identifying individuals at risk of depression on the basis of their social media content.

Performance metrics derived from the confusion matrix offer a thorough evaluation of the proposed model's efficacy in classifying depressions. The performance of the system is mainly evaluated on four parameters accuracy, precision, recall, F1-score. These measures, which are based on the concepts of False Positive (FP), False Negative (FN), True Negative (TN), and True Positive (TP), are essential for assessing the model's performance. The calculation of accuracy involves dividing the total number of predictions by the number of right predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

The exactness of a prediction is measured by its precision, or the number of true positives. Instead, recall quantifies completeness, or the number of real positives that were anticipated as positives.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1-Score = 2 * \left(\frac{Precision*Recall}{Precision+Recall}\right) \quad (13)$$

The performance metrics of the proposed system are detailed in Fig.18, illustrating its effectiveness in accurately predicting depression. The accuracy metric stands at 98.65%, indicating the overall correctness of the model's predictions compared to the total number of predictions made. Precision, measured at 99.13%, reflects the model's capability to correctly identify depression among those predicted. Recall, which measures at 98.09%, signifies the model's ability to accurately retrieve all instances from the dataset. The F1-score, calculated at 98.6%, harmonizes precision and recall into a single metric, offering a balanced assessment of the model's performance. These performance measures help assess the model's effectiveness in detecting depression and provide insights into its strengths and weaknesses.

Table 2. Performance comparison of the proposed model with other classifiers

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	85.23	86.21	83.45	84.80
Support Vector Machine	88.95	87.76	90.12	88.92
Random Forest	90.37	91.02	89.56	90.29
Standard LSTM	94.12	94.50	93.70	94.10
Proposed Bi-LSTM	**98.65**	**99.13**	**98.09**	**98.6**

The Receiver Operating Characteristic (ROC) curve serves as a visual representation of the effectiveness of a binary classification model. This graphical depiction showcases how the True Positive Rate (Sensitivity) relates to the False Positive Rate (calculated as 1 - Specificity) across different classification thresholds. The trade-off between correctly detecting positive instances (depressive tweets) and incorrectly classifying negative cases (non-depressive tweets) is represented by each point on the curve, which corresponds to a particular threshold value as shown in Fig. 19.

To ensure a fair and reproducible comparison of the proposed Bi-LSTM model with existing classifiers, rigorous experimentation was conducted on the Kaggle: Depression on Twitter dataset. By utilizing the same dataset for all classifiers, including Bi-LSTM, Logistic Regression, Support Vector Machines (SVM), Random Forest, and standard LSTM, the comparison was conducted under identical conditions, eliminating any inconsistencies that could arise from varying data sources.

The dataset preprocessing, including tokenization, vectorization, and normalization, was applied uniformly across all models. Additionally, each model was optimized using the same hyperparameters to ensure that the evaluation was conducted under identical conditions. The hyperparameters such as epochs, learning rate, and batch normalization were carefully tuned for each model to ensure optimal performance. This involved conducting grid searches or random searches for hyperparameter optimization, where multiple combinations of these parameters were tested for each classifier. The results of the different classifiers were then evaluated using the same performance metrics: accuracy, precision, recall, and F1-score, to ensure a comprehensive and comparable analysis.

The use of batch normalization was an important step in improving the stability and convergence of each model. It helped mitigate issues related to vanishing gradients and allowed for faster training, especially when using deep learning models like Bi-LSTM and LSTM. By normalizing the activations within each layer, batch normalization stabilized the learning process across all models. The learning rate was also adjusted for each model to avoid issues of underfitting or overfitting, ensuring that the models were neither too slow in converging nor prone to instability.

The number of epochs was another hyperparameter that was tuned to balance model performance. Too few epochs might result in underfitting, where the model fails to learn adequately from the data, while too many epochs could lead to overfitting, where the model memorizes the data without generalizing well to unseen instances.

Table 2 and Fig.20 presents the comparative performance of different classifiers on depression detection. Logistic Regression, with an accuracy of 85.23%, demonstrates the lowest performance across all metrics. Its precision (86.21%) and recall (83.45%) indicate that while the model performs reasonably well in predicting positive instances of depression, it still misses a significant portion of the actual depression cases, as reflected by its relatively low recall. This results in an F1-score of 84.80%, which balances precision and recall.

The Support Vector Machine (SVM) shows an improvement with an accuracy of 88.95%. It achieves a precision of 87.76%, indicating it correctly identifies depression cases with higher consistency than Logistic Regression. However, its recall is slightly higher at 90.12%, suggesting it is more adept at identifying actual depression cases but still exhibits some false positives. The F1-score of 88.92% balances the precision and recall, demonstrating the SVM's effective but imperfect performance.

Random Forest, another ensemble model, outperforms both Logistic Regression and SVM with an accuracy of 90.37%. Its precision (91.02%) and recall (89.56%) indicate a more balanced performance, as it is better at correctly identifying depression cases and retrieving them from the dataset. The F1-score of 90.29% confirms that Random Forest strikes a good balance between false positives and false negatives, making it more reliable than the previous models.

The Standard LSTM, which captures sequential dependencies, performs significantly better with an

accuracy of 94.12%. Its precision (94.50%) and recall (93.70%) indicate that it effectively identifies and retrieves depression cases with high accuracy. The F1-score of 94.10% reflects a well-rounded performance, though it still falls short of the Bi-LSTM model.

The proposed Bi-LSTM model stands out with the highest performance across all metrics. It achieves an impressive accuracy of 98.65%, significantly higher than any of the other models. With a precision of 99.13%, it is extremely accurate in predicting depression cases, minimizing false positives. Its recall of 98.09% ensures that almost all instances of depression are detected, reducing false negatives. The F1-score of 98.6% confirms the overall effectiveness of the Bi-LSTM model, providing a highly balanced approach to depression detection. The Bi-LSTM model's superior performance can be attributed to its bidirectional architecture, which allows it to learn both past and future contexts of the text, making it better equipped to understand the nuances in short, context-dependent text like social media posts. This enables the model to more accurately detect depression in tweets, outperforming the other models, including the standard LSTM.

5. Conclusion

The proposed study offers a comprehensive exploration of depression detection using advanced DL and NLP techniques. Through the evaluation of textual content via social media platforms, our proposed approach leverages SA and Bi-LSTM architecture to achieve impressive results, including an accuracy of 98.65%, precision of 99.13%, recall of 98.09%, and an F1 score of 98.6. The combination of sentiment analysis and Bi-LSTM proves to be a powerful tool for identifying nuanced emotional patterns associated with depression. The success of this model holds significant implications for early detection, intervention, and support for individuals at risk of depression. By harnessing the capabilities of deep learning, we contribute to the advancement of mental health assessment methodologies and offer a promising direction for future research. The demonstrated accuracy and effectiveness of our approach underscore the potential of leveraging advanced machine learning techniques to address complex societal issues, particularly in the realm of mental health. As technology continues to evolve, our model stands as a testament to the positive impact that data-driven innovation can have on improving the well-being of individuals worldwide.

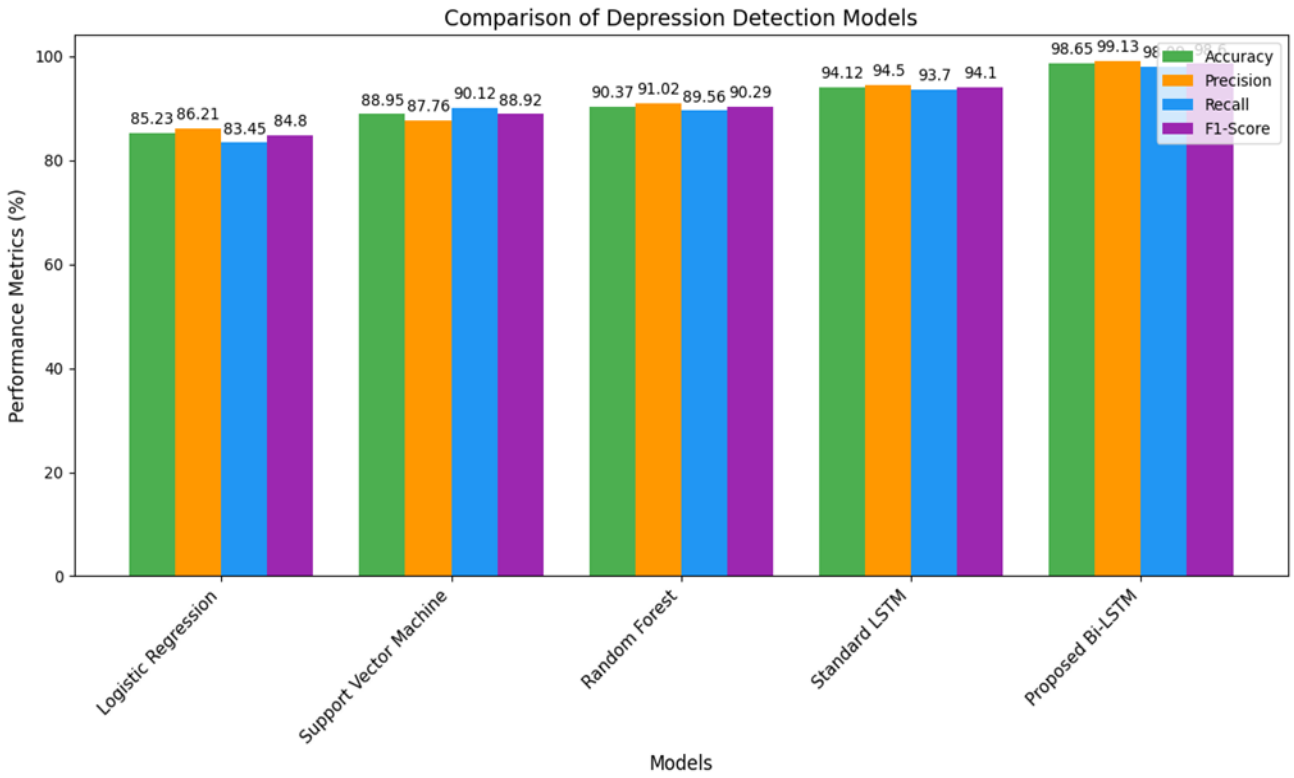

Figure.20 Performance comparison of the proposed model with other classifiers

Notations:

x_t	Input at time step t
h_{t-1}	Hidden state from the previous time step
C_{t-1}	Memory cell from the previous time step
W_f	Weight matrix for the forget gate
W_i	Weight matrix for the input gate
W_C	Weight matrix for the candidate memory cell update
W_o	Weight matrix for the output gate
b_f	Bias term for the forget gate
$b_{i,}$	Bias term for the input gate
b_c	Bias term for the candidate memory cell update
b_o	Bias term for the output gate
f_t	Forget gate activation at time step t
i_t	Input gate activation at time step t
C_t	Memory cell at time step t
C_t	Output gate activation at time step t
E_{w_i}	Embedding vector for the i^{th} word

Acknowledgments

I would like to express my sincere gratitude to all those who contributed to the completion of this research paper. I extend my heartfelt thanks to family, colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

References

[1] U.S. Department of Health and Human Services, *Healthy People 2010: Understanding and Improving Health*, Washington, DC, Government Printing Office, 2000.

[2] D. V. Vigo, A. E. Kazdin, N. A. Sampson, I. Hwang, J. Alonso, L. H. Andrade, ... and R. C. Kessler, "Determinants of effective treatment coverage for major depressive disorder in the WHO World Mental Health Surveys", *Int. J. Mental Health Syst.*, Vol. 16, No. 1, pp. 29, 2022.

[3] M. A. Torres, "What in the World is Wrong With You? Rethinking Depression and Suicide Among 18-25-Year-Old Sailors", *Doctoral dissertation, SPIRITUAL HEALTH*, 2024.

[4] R. T. Mok, R. Sing, X. Jiang, and S. See, "Investigation of Social Media on Depression", In: *Proc. 9th Int. Symp. Chinese Spoken Lang. Process. (ISCSLP)*, pp. 488–491, 2014.

[5] S. Rodrigues, B. Bokhour, N. Mueller, N. Dell, P. E. Osei-Bonsu, S. Zhao, M. Glickman, S. V. Eisen, and A. R. Elwy, "Impact of stigma on veteran treatment seeking for depression", *Am. J. Psychiatr. Rehabil.*, Vol. 17, No. 2, pp. 128-146, 2014.

[6] K. Sibelius, "Increasing access to mental health services", *White House Blog*, Apr. 2013. [Online]. Available: http://www.whitehouse.gov/blog/2013/04/10/increasingaccess-mentalhealthservices.

[7] S. Aslam, "Twitter by the numbers: Stats, demographics & fun facts", *Omnicore Agency*, 2018.

[8] J. Clement, "Twitter: number of monthly active users 2010–2019", *Statista*, 2019. [Online]. Available: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[9] F. Benamara, V. Moriceau, J. Mothe, F. Ramiandrisoa, and Z. He, "Automatic Detection of Depressive Users in Social Media", In: *Proc. of Conférence francophone en Recherche d'Information et Applications (CORIA)*, Rennes, France, pp. 16-18 2018.

[10] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media", In: *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, New York, 2013.

[11] A. Kumar, A. Sharma, and A. Arora, "Anxious depression prediction in real-time social data", In: *Proc. of International Conference on Advanced Engineering, Science, Management and Technology (ICAESMT)*, 2019.

[12] L. P. Richardson et al., "Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents", *Pediatrics*, Vol. 126, pp. 1117-1123, 2010.

[13] G. Geetha et al., "Early Detection of Depression from Social Media Data Using Machine Learning Algorithms", In: *Proc. of 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, pp. 1-6, 2020, doi: 10.1109/ICPECTS49113.2020.9336974.

[14] X. Wang, C. Zhang, and L. Sun, "An improved model for depression detection in micro-blog social network", In: *Proc. of 2013 IEEE 13th*

International Conference on Data Mining Workshops, Dallas, TX, USA, pp. 80-87, 2013.

[15] D. A. Musleh et al., "Twitter Arabic Sentiment Analysis to Detect Depression Using Machine Learning", *Computers, Materials & Continua*, Vol. 71, No. 2, pp. 1301-1313, 2022.

[16] P. Kumar et al., "Feature based depression detection from Twitter data using machine learning techniques", *Journal of Scientific Research*, Vol. 66, No. 2, pp. 220-228, 2022.

[17] K. A. Govindasamy and N. Palanichamy, "Depression detection using machine learning techniques on Twitter data", In: *Proc. of 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 960-966, 2021.

[18] H. S. AlSagri and M. Ykhlef, "Machine learning-based approach for depression detection in Twitter using content and activity features", *IEICE Trans. Inf. Syst.*, Vol. 103, No. 8, pp. 1825-1832, 2020.

[19] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM", *Multimedia Tools and Applications*, Vol. 81, No. 17, pp. 23649-23685, 2022.

[20] Y. Chen et al., "Sentiment Analysis Based on Deep Learning and Its Application in Screening for Perinatal Depression", In: *Proc. of 2018 IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, pp. 451-456, 2018.

[21] D. S. Khafaga et al., "Deep Learning for Depression Detection Using Twitter Data", *Intelligent Automation and Soft Computing*, Vol. 36, No. 2, pp. 1301-1313, 2023.

[22] A. Amanat et al., "Deep learning for depression detection from textual data", *Electronics*, Vol. 11, No. 5, pp. 676, 2022.

[23] V. Tejaswini, K. S. Babu, and B. Sahoo, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model", *ACM Trans. Asian and Low-Resource Lang. Inf. Process.*, 2022.

[24] E. P. Sri, K. S. Savita, and M. Zaffar, "Depression Detection in Tweets from Urban Cities of Malaysia using Deep Learning", In: *Proc of 2021 7th International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp. 1-6, 2021.

[25] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts", *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, Vol. 4, No. 3, pp. 375-380, 2019, doi: 10.22219/kinetik.v4i4.912.

[26] G. Singh et al., "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification", In: *Proc. of 2019 Int. Conf. Autom. Comput. Technol. Manag. (ICACTM)*, pp. 593-596, 2019, doi: 10.1109/ICACTM.2019.8776800.

[27] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, No. 3, pp. 588–601, 2018.