DETECTING EMOTIONAL STATE OF DEPRESSION IN SOCIAL MEDIA POSTS USING LOGISTIC REGRESSION-RECURSIVE FEATURE ELIMINATION

Wang Li[1] , Wandeep Kaur [2*], Chen Wangmei[3]

[1, 2, 3]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia,
43600 Selangor, Malaysia

Emails: wangliwangli917@gmail.com[1], wandeep@ukm.edu.my[2*], chancandy898@gmail.com[3]

ABSTRACT

Depression detection through social media has garnered widespread attention due to its potential for early intervention in mental health issues. This study aims to detect depressive users based on their content shared on social media using machine learning techniques. Given the complexity and diversity of depressive text, existing research still falls short in exploring comprehensive feature extraction techniques. To address this challenge, this study proposes an integrated framework for detecting depressive tendencies through multi-dimensional feature extraction and selection techniques. The proposed approach combines TF-IDF with N-grams, DistilBERT embeddings, and SentiWordNet to capture linguistic, semantic, and emotional features. Additionally, logistic regression-based recursive feature elimination (LR-RFE) is employed to optimize high-dimensional feature sets by reducing redundancy and emphasizing key indicators.Experiments conducted on the CLEF eRisk dataset revealed varying levels of effectiveness across individual feature extraction methods. Notably, multi-feature integration significantly enhanced classification performance, achieving an accuracy of 80.8% and an F1 score of 80.54% with the combined feature set. Feature selection further improved model efficiency and performance. These findings contribute to advancing automated depression detection and lay a foundation for developing scalable and interpretable machine learning models for mental health assessment.

Keywords: Machine learning; Feature extraction; Feature selection; Depression detection; Social media.

1.0 INTRODUCTION

Depression is a widespread mental health disorder affecting up to 350 million people globally, with its incidence continuing to rise[1], [2]. Depression's primary symptoms manifest in language, emotions, and behavioral patterns, which are frequently characterized by increased self-focus, heightened emotional distress, and decreased social interaction[3]. Individuals suffering from depression, for example, may experience frequent feelings of sadness, frustration, or tears, lose interest in most activities, become fatigued, and have difficulty thinking or concentrating[4], [5], [6]. These symptoms have a significant impact on an individual's well-being and may even lead to suicidal thoughts, emphasizing the importance of prompt diagnosis and treatment. Studies have shown that early detection of depression can significantly reduce its negative effects[7], [8]. However, despite advances in medical and technological fields, early detection and intervention remain difficult, owing to people's lack of self-awareness and the limitations of traditional diagnostic methods[9], [10].

In recent years, social media platforms such as Twitter, Facebook, and Reddit have become import channels for individuals to express their emotions and share psychological states in real-time, offering new opportunities for early depression detection. These platforms provide a rich, unfiltered dataset for analyzing emotional states, enabling researchers to uncover linguistic markers associated with depression[3],[11]. Previous research has found that people who are depressed have distinct language patterns, such as increased use of first-person singular pronouns, less lexical diversity, and a higher frequency of negative expressions related to sadness and hopelessness[12],[13]. These linguistic patterns demonstrate the multifaceted nature of depression-related text, which includes a complex interplay of lexical, semantic, and emotional features. As a result, an increasing number of studies have explored the use of machine learning (ML) and natural language processing (NLP) techniques to analyze these features for automatic depression detection, including lexical features (such as TF-IDF and N-gram), semantic features (like BERT and Word2Vec), and emotional features (such as sentiment analysis and sentiment lexicons) [9],[14],[15],[16]. However, due to the complexity of these linguistic patterns, relying on single-feature analysis often fails to accurately detect depression[17], [18]. Previous studies have explored various combinations of features. For instance, Muzafar et al.[19]

combined BoW, TF-IDF, and Word2Vec to capture lexical and semantic features; Noor et al.[20] examined LIWC, N-gram, and vector embeddings to focus on psychological and linguistic features; and Eichstaedt et al.[21] integrated LDA, N-gram, and LIWC to capture topic modeling and psychological indicators. However, the integration of multiple feature combinations still needs further exploration. To address these challenges, this study proposes a multi-dimensional feature extraction framework for detecting depression. The proposed approach integrates Term Frequency-Inverse Document Frequency (TF-IDF) with N-grams for lexical analysis[17],[22], DistilBERT embeddings for contextual semantic understanding[23], and SentiWordNet for sentiment polarity analysis[24], thus providing a more comprehensive representation of linguistic patterns in depression-related text. Furthermore, logistic regression-based recursive feature elimination (LR-RFE) is employed to optimize these features by reducing redundancy while retaining critical indicators of depression[25]. The main contributions of this study are as follows:

1. Multi-Dimensional Feature Extraction: The integration of TF-IDF, N-grams, DistilBERT embeddings, and SentiWordNet captures the multifaceted nature of depressive language, enabling robust representation across lexical, semantic, and emotional dimensions.
2. Optimized Feature Selection: By applying LR-RFE, this study ensures efficient feature selection, reducing noise and redundancy while maintaining interpretability.
3. Empirical Validation: The proposed methodology is evaluated on the CLEF eRisk dataset, demonstrating significant improvements in classification accuracy, robustness, and efficiency.

Experimental results validate the effectiveness of this multi-dimensional approach, highlighting its ability to enhance classification performance while balancing model complexity and stability. This research not only provides a practical framework for depression detection but also lays the groundwork for future advancements in mental health analytics using machine learning. The remainder of this paper is organized as follows: Section 2 reviews the related work on depression detection. Section 3 provides a detailed explanation of the methodology. Section 4 presents the experimental results, analysis, and discussion. Finally, Section 5 concludes the paper and outlines future work.

2.0 LITERATURE REVIEW

The detection of depression through machine learning (ML) has seen significant advancements, utilizing social media as a valuable data source to uncover linguistic, emotional, and psychological patterns. Various studies have demonstrated the efficacy of ML techniques in classifying depressive text, Table 1 shows the recent related studies. For instance, experiments by Chiong et al.[15] demonstrated that using multiple machine learning methods such as svm, LR, and adaptive boosting is effective even when training on text that does not contain the words "depression" or "diagnosis". Samanvitha et al.[26] achieved the highest accuracy of 83% using NB algorithm on the eRisk dataset. Similarly, Sharma et al.[27] applied a hybrid LSTM-CNN model to analyze depressive language patterns, achieving an impressive accuracy of 97%. These studies highlight the potential of ML in automating depression detection by leveraging large-scale user-generated content. However, they also reveal a common limitation: the reliance on specific algorithms or datasets often overlooks the deeper semantic and emotional nuances inherent in depressive text.

Feature extraction plays a pivotal role in addressing these limitations, as it transforms unstructured textual data into structured representations that capture linguistic patterns and emotional cues. Commonly used techniques include Term Frequency-Inverse Document Frequency (TF-IDF), N-grams, BOW and Word2Vec embeddings[2],[22],[28]. Muzafar et al.[19]demonstrated that integrating Word2Vec with SVM achieved an accuracy of 81.79%, highlighting the importance of capturing semantic relationships between words. Similarly, Tadesse et al. [29] combined TF-IDF and N-grams with multiple ML algorithms, achieving an F1 score of 0.93. Despite their utility, these methods often fail to represent the deeper contextual semantics and emotional complexities of depressive text. For example, Stankevich et al.[30] utilized the CLEF/eRisk 2017 dataset and revealed that features combining TF-IDF and morphological sets achieved an F1 score of 63%, underscoring the need for more comprehensive approaches.

To address these challenges, recent studies have explored integrating complementary feature extraction methods to enhance classification performance. For instance, Hosseini-Saravani et al.[31] introduced a bipolar feature vector to capture psychological characteristics of depressed individuals, achieving an F1 score of 82.75%. While effective, their approach relied heavily on predefined features, limiting its ability to capture more nuanced semantic and emotional patterns inherent in depressive text. Muzafar et al. [19] further explored combinations of BOW, TF-IDF, and Word2Vec with RF and SVM, validating the effectiveness of multi-feature integration. However, these methods still lack the ability to fully utilize advanced contextual embeddings that can provide richer semantic representations Advanced techniques such as SentiWordNet and BERT have also shown promise in capturing emotional polarity and semantic richness. For example, Chiong et al.[32] achieved an accuracy of 92.61% by integrating SentiWordNet with contextual models, underscoring the utility of sentiment analysis tools. Nevertheless, their reliance on specific sentiment lexicons limits the generalizability of the approach to domains with more complex emotional expressions. Similarly, Ansari et al.[33] demonstrated that combining lexicon-based approaches with LSTM models improved classification accuracy to 75%, but the method's dependence on lexicon-driven features may restrict its adaptability to diverse datasets. These studies highlight the importance of leveraging diverse linguistic and emotional features to enhance the robustness and interpretability of ML models in depression detection.

While feature extraction provides critical insights into depressive text, high-dimensional feature spaces pose challenges such as noise and redundancy, which can hinder model efficiency and interpretability[18]. Feature selection techniques, such as logistic regression-based recursive feature elimination (LR-RFE), have been widely adopted to address these issues[34]. Ansari et al. [33]showed that LR-RFE improved classification accuracy from 64% to 75%, while Liu et al.[18] combined RFE with ensemble models, achieving 95.63% accuracy on balanced datasets. These techniques not only refine feature sets but also reduce computational complexity, making them particularly valuable for analyzing complex depressive text.

Table 1: Recent research on depression detection

Authors	Dataset	Methods	Feature extraction	Best Results	Limitations
Fang et al. (2022)	Reddit	SVM, NB	a)TF-IDF;b) LIWC; c)Word2Vec;d) TF-IDF & Word2Vec.	Accu: 95.68% F1: 92.84	Limited exploration of high-dimensional data
Ansari et al. (2023)	CLPsych2015, Reddit, eRisk	LSTM, Ensemble,Hybrid Lexicon-Based LR	Sentiment Features, AFINN,NRC(NRC_SA), MPQA, and SenticNet	F1:77%	Dependency on sentiment lexicons
Noor et al. (2023)	Twitter	NB,DT, SVM, Kneighbors Classifier, RF, LR, Bagging Model	LIWC,N-gram, Vectors embedding	Accu: 98.33% F1: 92.15%	Over-reliance on LIWC
Chiong et al. (2021)	Twitter	LR,SVM, MLP, DT, RF, AB,BP,GB	BOW,Count vectorisation,N-gram words	Accu: 92.61%	Limited use of deep contextual embeddings
Muzafar et al. (2023)	Twitter	LSTM,NB, SVM, RF	BOW,TF-IDF, Count Vectorization,Word2Vec	Accu: 81.79%	No advanced contextual embeddings
Stankevich et al. (2018)	eRisk	RF, SVM	Stylometric features,Morphology features, TF-IDF, Word embedding, Bi-gram	F1: 0.63	Basic feature extraction methods
Liu et al. (2022)	Weibo	NB,KNN, SVM,Regression,Stacking ensemble	Dictionary,Post behaviour information tag	Accu: 95.63%	Over-reliance on dictionary-based features
Hosseini-Saravani et al. (2020)	eRisk	NB	Bipolar feature extraction	F1: 82.75%	Simple Bayesian model
Eichstaedt et al. (2018)	Facebook	LR	LDA, N-gram, LIWC	Accu: 72%	Limits semantic understanding
Janatdoust et al. (2022)	Reddit	Ensemble model	BERT	Accu: 61% F1: 54%	Limiting understanding of features
Tadesse et al. (2019)	Reddit	LR, SVM, Ada boost, RF, MLP	N-gram, LIWC, LDA	F1:93%	Focus on basic features limits nuance

Despite these advancements, the integration of multi-dimensional feature extraction techniques with feature selection methods remains underexplored. As emphasized by [18],[35], combining diverse feature extraction methods offers a promising approach to capturing the intricate patterns of depressive text. Building on these insights, this study proposes a novel framework that integrates TF-IDF, N-grams, SentiWordNet, and DistilBERT to provide a comprehensive representation of linguistic, semantic, and emotional features. By employing LR-RFE for optimized feature selection, this approach ensures robust representation, balancing accuracy, interpretability, and computational efficiency, thereby providing a practical framework for automated depression detection.

3.0 METHODOLOGY

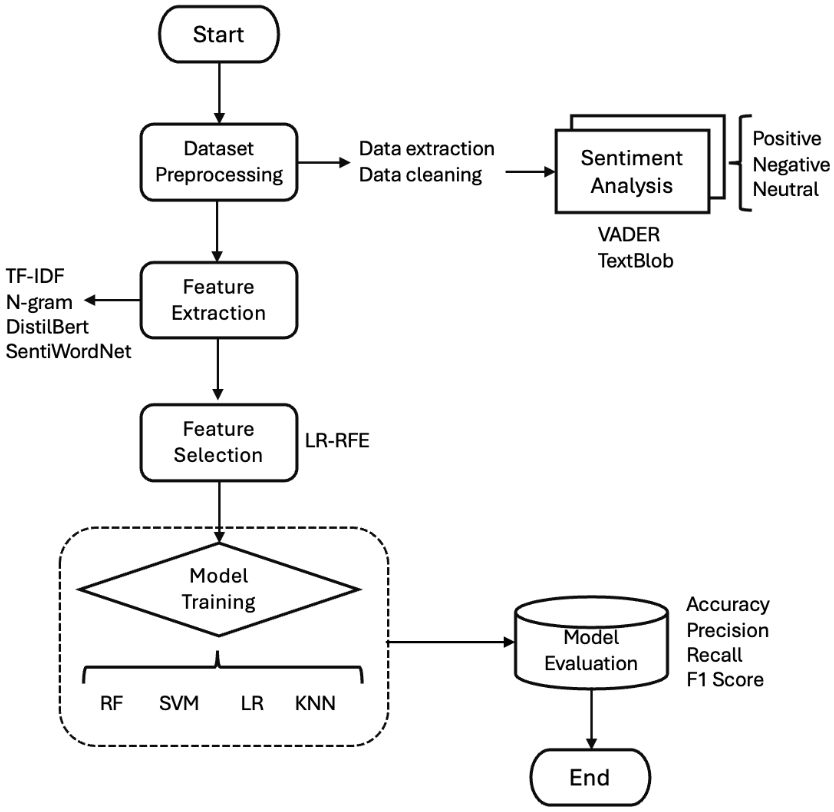

Fig. 1: The medthodology of Depression Detection

The flowchart shown in Fig. 1 is the process of the method in this study, which can intuitively show the overall method of this study. The main task of this study is to explore the effectiveness of combining different feature extraction and selection methods to detect depression in social media text data.

3.1 Data Preprocessing

This study used the CLEF eRisk 2017 dataset, which consists of user posts from the social media platform Reddit that have been labeled to distinguish between depressive and non-depressive users[40], [41], [42]. The dataset is designed to identify mental health-related issues such as depression, self-harm, pathological gambling, and eating disorders, providing a valuable textual resource for early mental health detection. The eRisk dataset contains over 2,000 files, with each file including elements such as TITLE, DATE, and TEXT (as shown in Fig. 2). The TITLE represents the heading of the user's post, while the DATE indicates the timestamp of when the post was made. The TEXT is the body of the post, containing the content provided by the user. INFO includes metadata about the post, such as the user's ID or any additional tags. In this study, only TEXT variables were used to detect depressive states because they are the most closely related to the textual content of the posts. To achieve the study's objective of detecting depressive tendencies in textual data, 1,079 text records were extracted from the posts of a user labeled as depressive. To ensure the quality and relevance of the dataset, the text underwent preprocessing steps such as stopword removal, elimination of non-English characters, case normalization, and tokenization. Stopwords were removed using the NLTK library, and text was tokenized into unigrams and bigrams using a custom tokenizer. Additionally, negation words, which play a critical role in altering sentence polarity, and short posts, which often contain valuable insights into the user's emotional state, were retained during preprocessing.


```
<INDIVIDUAL>
  <ID>subject1</ID>
  <WRITING>
    <TITLE></TITLE>
    <DATE>2017-08-18 11:34:09</DATE>
    <TEXT>Vulcan's ultimate landing at max range is so satisfying :3. BOOM!</TEXT>
    <INFO>Reddit post</INFO>
  </WRITING>
  <WRITING>
    <TITLE>Defense item in mid lane?</TITLE>
    <DATE>2017-08-20 15:26:34</DATE>
    <TEXT>Is there any defensive item (physical) viable to use in mid lane? I'm having trouble to choose a defensive item when i'm playing
against certain picks like Bastet or Susano. Back in the day we had Celestial Legion Helm but now i think it's not worth it. Dynasty plate
helm has been recently nerfed so i don't know what to build... maybe breastblate for cooldown?</TEXT>
    <INFO>Reddit post</INFO>
  </WRITING>
  <WRITING>
    <TITLE></TITLE>
    <DATE>2017-08-20 15:47:00</DATE>
    <TEXT>Is it still op? His new passive is a little bit strange</TEXT>
    <INFO>Reddit post</INFO>
  </WRITING>
```


Fig. 2: The raw dataset of the eRisk

A frequency distribution analysis (Fig. 3) revealed that depressive users frequently employ first-person pronouns and negative-emotion words. These patterns reflect heightened self-focus and emotional distress, characteristics that align closely with the psychological traits of depression.

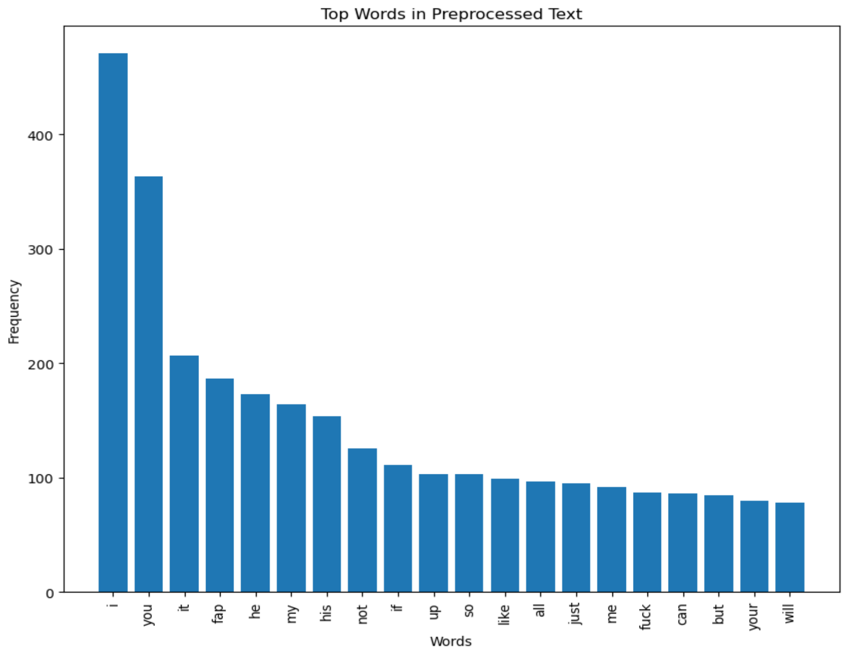

Fig. 3: Pre-processed Word Frequency

3.2 Sentiment Analysis

Sentiment analysis is another crucial component of this study, aimed at identifying and categorizing the emotions expressed in text. To address the lack of labels in the dataset, this study utilized two widely adopted sentiment analysis tools to classify the emotional polarity (positive, neutral, or negative) of the posts: VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob[43].

VADER is specifically designed for social media content, assigning sentiment scores using a predefined lexicon while capturing nuances such as intensity modifiers and punctuation emphasis[44]. In contrast, TextBlob employs a rule-based framework to evaluate polarity and subjectivity, but it tends to classify subtle emotional content as neutral[45]. Comparative results (as shown in Fig. 4) demonstrate noticeable differences between VADER and TextBlob. For example, VADER identified 334 negative texts, compared to 213 by TextBlob, likely due to VADER's sensitivity to emotionally expressive elements such as slang and punctuation. Conversely, TextBlob classified more texts as neutral (396) than VADER, reflecting its conservative approach, which may overlook subtle emotional cues, particularly in texts with complex emotional undertones. Both tools performed similarly in detecting positive emotions, with VADER identifying 289 positive texts and TextBlob detecting 306. Fig. 5 and Fig. 6 illustrate the sentiment word clouds generated by TextBlob and VADER, respectively. In these visualizations, warm colors, green, and blue represent positive, neutral, and negative sentiments, with font size indicating keyword frequency. These visualizations further underscore the comparative analysis, highlighting that while both tools effectively recognize positive emotions, VADER's superior ability to capture the intensity and nuances of negative emotions makes it particularly well-suited for analyzing depression-related content. Based on these observations and to ensure consistency in data labeling and sensitivity to emotional expressions, VADER's sentiment labels were selected as the primary labels for this study.

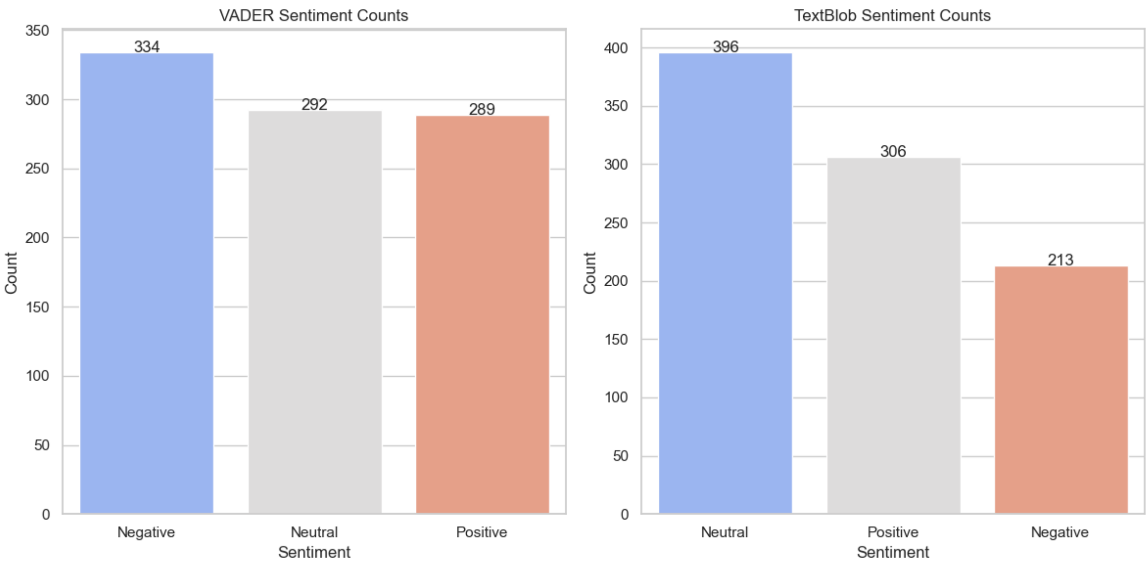

Fig. 4: Sentiment counts for VADER and TextBlob

Fig. 5: TextBlob Sentiment WordCloud

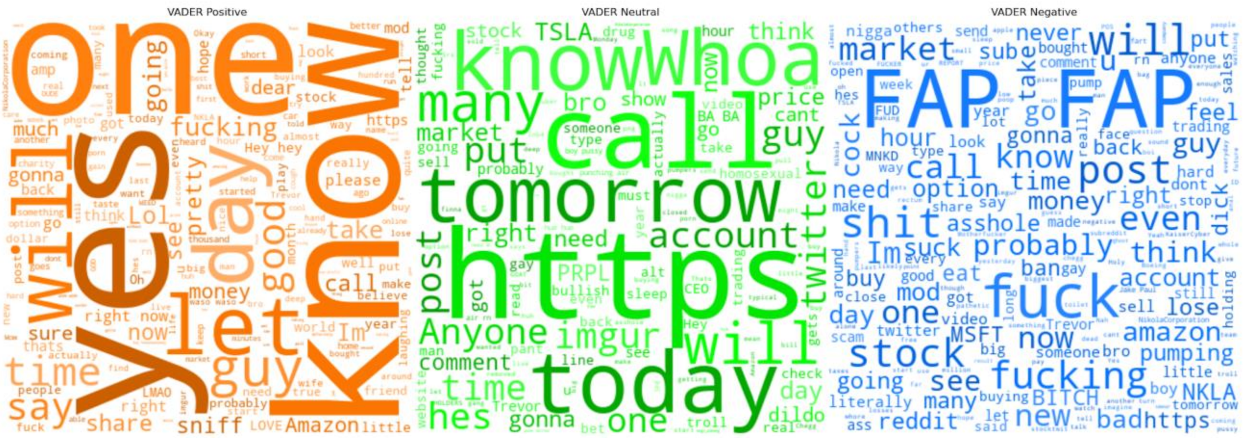

Fig. 6: VADER Sentiment WordCloud

3.3 Feature Extraction

Feature extraction is a fundamental process that transforms unstructured textual data into structured formats suitable for machine learning algorithms, enabling the identification of patterns indicative of depression in social media posts[28]. This study employed three distinct approaches—TF-IDF with N-grams[22], DistilBERT[46], and SentiWordNet[47]—to capture various linguistic, semantic, and emotional features from the dataset.

TF-IDF was selected for its ability to capture term importance and contextual relationships within the text[48]. By calculating term frequency (TF) and inverse document frequency (IDF), this approach quantifies the weight of each word, reflecting its relevance within a document and its rarity across the corpus. N-grams, encompassing unigrams, bigrams, and trigrams, were incorporated to capture local linguistic patterns. These patterns are particularly effective in identifying emotionally significant phrases, such as "I feel sad" or "deeply distressed." The resulting TF-IDF scores generate a sparse feature matrix, ensuring comprehensive lexical representation while accommodating variations in word usage and context.

DistilBERT, as a simplified version of BERT, offers several advantages for depression detection tasks. Its lightweight architecture significantly reduces computational costs and training time, while retaining the ability to capture intricate contextual nuances in text. This makes it particularly well-suited for analyzing the complex semantics inherent in depressive language[49]. A comparative study by Cortiz et al.[46] highlighted that DistilBERT achieves competitive performance compared to full BERT models, with only marginal reductions in accuracy but substantial improvements in efficiency. Such characteristics are crucial in depression detection, where processing large-scale social media data efficiently is often a key requirement. Building on these advantages, this study applied DistilBERT as follows: Each text sequence was preprocessed using special tokens ([CLS] and [SEP]) and converted into numerical token IDs. The preprocessed sequences were then processed through the DistilBERT model, which produced 768-dimensional feature vectors representing the semantic richness of each text fragment, effectively encapsulate contextual subtleties..

Finally, SentiWordNet, a lexicon-based sentiment analysis tool, was used to evaluate words based on their objectivity, positivity, and negativity[47]. SentiWordNet was chosen for its focus on emotional polarity, complementing the semantic depth captured by DistilBERT. Leveraging WordNet's lexical database and the Lesk algorithm for word sense disambiguation, SentiWordNet assigns sentiment scores to individual words and aggregates them at the sentence level. This process provides a structured framework for evaluating the emotional tone of text, effectively highlighting depressive sentiments. For example, a sentence containing the word "sad" may exhibit a high negative score with lower positive and neutral scores, emphasizing its depressive undertone.

The feature extraction process employed in this study, as illustrated in Algorithm1, demonstrates the complementary strengths of the three methods. TF-IDF generates the broadest feature set, capturing linguistic patterns but incurring higher computational costs. DistilBERT offers moderate feature dimensionality while excelling in semantic representation. SentiWordNet provides low-dimensional features with high computational efficiency, effectively emphasizing emotional content. Integrating these methods into a unified workflow ensures multi-dimensional feature capture, encompassing linguistic patterns, semantic depth, and emotional context.

Algorithm 1 Process for Feature Extraction

Data: Preprocessed dataset

Output: Individual feature sets from TF-IDF, DistilBERT, and SentiWordNet

for all user $\in$ dataset **do**

TF-IDF with N-gram:

Extract unigrams, bigrams, and trigrams from text.

Compute Term Frequency-Inverse Document Frequency (TF-IDF) scores for each N-gram:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \cdot \text{IDF}(t)$$

Generate a sparse matrix of up to 3,000 features.

DistilBERT for Semantic Features:

Tokenize the text using DistilBERT's tokenizer.

Add special tokens such as [CLS] and [SEP].

Convert tokens into numeric IDs and process through DistilBERT to extract 768-dimensional sentence embeddings.

SentiWordNet for Sentiment Features:

Tokenize text and assign part-of-speech tags using SpaCy.

Map words to synsets using WordNet.

Retrieve sentiment scores (positive, negative, neutral) for each word and aggregate scores at the sentence level.

end for

Algorithm 1: The process of feature extraction

3.4 Experimental Design

To evaluate the effectiveness of feature combinations in depression detection tasks, this study designed three experiments to assess the performance of individual features, the enhancement provided by feature combinations, and the performance improvements achieved through optimized feature selection(as shown in Fig. 7). These experiments were designed to explore the diversity and complementarity of feature extraction methods and validate their applicability to the classification of depressive text.

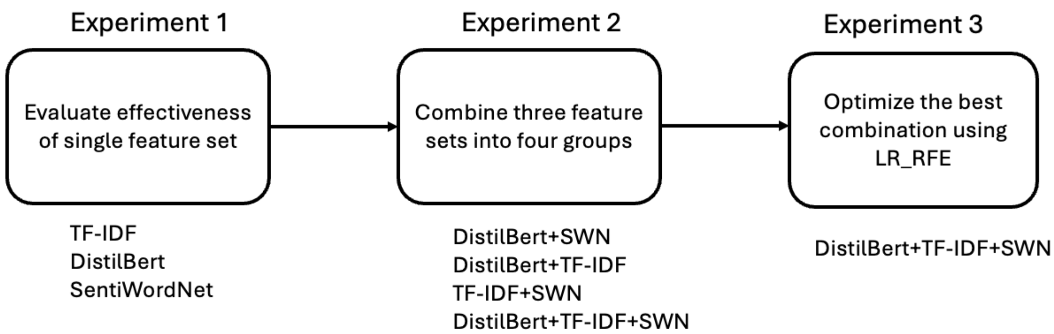

Fig. 7: The process of the three experiments

The first experiment aimed to evaluate the effectiveness of individual feature extraction methods. Each feature set—TF-IDF, DistilBERT, and SentiWordNet—was independently applied to train machine learning models, and standard classification metrics (accuracy, precision, recall, and F1 score) were used to assess performance. This phase provided a foundational understanding of the contribution of each feature set to depression detection. The results from this experiment served as a benchmark for evaluating the effectiveness of feature combinations.

The second experiment investigated the impact of combining multiple feature sets on model performance. Pairwise combinations and the integration of all three feature extraction methods were tested to explore their complementarity. Specifically, combinations such as BERT + SWN, TF-IDF + BERT, TF-IDF + SWN, and BERT + TF-IDF + SWN were created by horizontally concatenating feature vectors. These combinations aimed to integrate semantic richness, lexical relevance, and emotional polarity into a unified representation. This experiment sought to assess whether combining complementary features could enhance the model's ability to detect nuanced patterns and improve classification robustness.

The final experiment focused on optimizing the combined feature set through logistic regression-based recursive feature elimination (LR-RFE)[50]. This approach aimed to reduce dimensionality by systematically eliminating the least informative features while retaining the most relevant ones. The process started with a comprehensive feature set that included all features from TF-IDF, DistilBERT, and SentiWordNet. A logistic regression model was trained, and RFE was applied iteratively to identify the optimal subset of features for maximizing classification performance. The number of features was adjusted incrementally to evaluate their impact on the model's performance, with the ultimate goal of identifying an optimal feature count that balanced dimensionality and model accuracy. The detailed results of this experiment are discussed in the results analysis section.

3.5 Model Selection

This study employed four commonly used machine learning models—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN)—to evaluate the effectiveness of feature extraction methods[29],[50],[1],[51]. The experimental design prioritized validating the effectiveness of the proposed feature extraction methods and their combinations in depression detection tasks. To maintain the generality and reproducibility of the study, model parameters were not further optimized. The selection of the four models was based on their widespread application in text classification tasks and their compatibility with diverse feature types, ensuring a comprehensive evaluation of the proposed methodologies.

- SVM[52]: Known for its ability to handle high-dimensional data and non-linear classification problems, SVM uses kernel functions to maximize the margin between classes, which is particularly useful for datasets with complex feature spaces, such as TF-IDF and semantic embeddings[53].

- RF[54]: As an ensemble model, RF combines multiple decision trees, capturing complex interactions among features while maintaining resistance to overfitting. It is particularly adept at leveraging feature importance mechanisms to highlight key contributors to classification tasks[55].

- KNN[56]: This model, based on proximity in feature space, is straightforward to implement and excels in small-sample scenarios. It effectively identifies local patterns and subtle variations in emotional expression in social media text[57].

- LR[58]: Renowned for its interpretability and computational efficiency, LR applies a sigmoid function to calculate probabilities, enabling it to predict the likelihood of a post expressing depressive sentiments. Its adaptability across diverse feature sets makes it a versatile choice for this task[59].

The models were implemented with default hyperparameters to ensure a focus on the impact of feature extraction and combinations. This strategy emphasizes the role of feature design in influencing performance, providing a foundation for future model optimization.

3.6 Performance Evaluation

This study employed four key evaluation metrics—accuracy, precision, recall, and F1 score—to assess the models' ability to accurately classify depressive and non-depressive posts. These metrics collectively ensure that the models are thoroughly evaluated, balancing their general accuracy with their ability to specifically detect depressive tendencies while avoiding misclassification[60],[61].

4.0 RESULTS AND ANALYSIS

The experimental results validate the performance differences among various feature extraction methods and classification models in the context of depression detection. Additionally, they highlight the critical role of multi-feature integration and feature selection in optimizing classification performance.

4.1 The Experimental Results Of Single Feature

Table 2: Results of single feature

Feature Set	Model	Accuracy	Precision	Recall	F1-Score
DistilBert	RF	0.648	0.6496	0.648	0.6446
	LR	**0.744**	**0.744**	**0.744**	**0.744**
	KNN	0.664	0.6696	0.664	0.6581
	SVM	0.536	0.5953	0.536	0.4143
TF-IDF	**RF**	**0.664**	**0.7069**	**0.664**	**0.6395**
	LR	0.616	0.6186	0.616	0.6093
	KNN	0.616	0.6554	0.616	0.5799
	SVM	0.552	0.5534	0.552	0.5521
SentiWordNet	**RF**	**0.68**	**0.6804**	**0.68**	**0.6786**
	LR	0.648	0.668	0.648	0.6314
	KNN	0.624	0.6296	0.624	0.6147
	SVM	0.64	0.6783	0.64	0.612

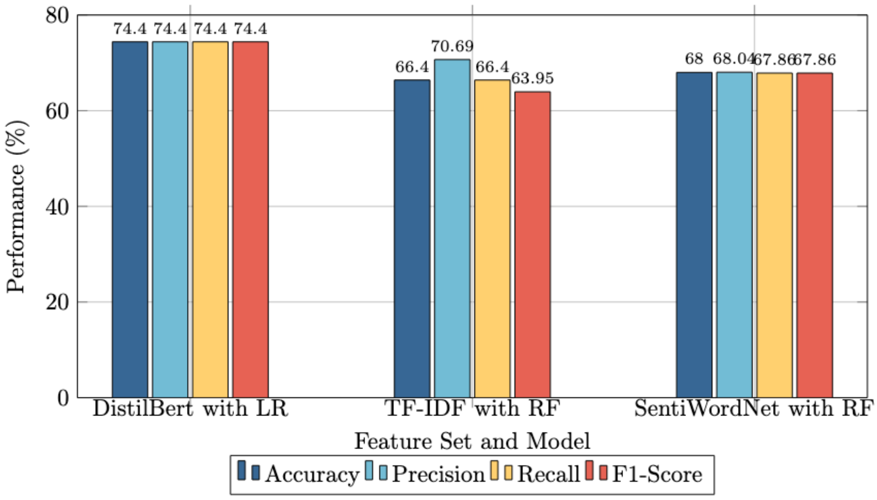

Fig. 8: The best results of single feature

Table 2 and Fig. 8 show all the results and the best result in the single-feature experiment. The single-feature experiments revealed notable differences in the performance of various feature extraction methods when paired with different classification models, providing insights into their compatibility and practical implications.

DistilBERT combined with Logistic Regression (LR) achieved the best overall performance, with an accuracy and F1 score of 74.4%. This result highlights the ability of DistilBERT embeddings to capture deep semantic patterns and contextual nuances in depressive language. The high F1 score suggests that this combination effectively balances precision and recall, making it particularly suitable for early depression detection, where false negatives (undetected depressive cases) can have severe consequences. In contrast, SVM underperformed with DistilBERT, yielding an F1 score of only 41.43%. This suggests that SVM struggled with high-dimensional semantic embeddings, often misclassifying neutral posts as depressive due to its limited adaptability to complex feature spaces.

Random Forest (RF) showed strong performance with TF-IDF features, achieving an accuracy of 66.4% and an F1 score of 63.95%. This indicates RF's robustness in handling high-dimensional sparse features, which are characteristic of TF-IDF representations. However, LR paired with TF-IDF yielded lower results (F1 score 60.93%), likely due to its reliance on linear relationships, which limits its ability to leverage sparse lexical patterns effectively.

SentiWordNet (SWN) features also performed well, particularly with RF achieving an accuracy of 68.0% and an F1 score of 67.86%. This demonstrates the value of emotional polarity in identifying depressive tendencies, as these features capture subtle shifts in sentiment often associated with depression. However, traditional models like KNN and SVM showed consistently poor performance across all single-feature experiments. KNN's reliance on proximity-based decision-making led to inconsistent results, while SVM's difficulty in high-dimensional feature spaces resulted in frequent misclassification of neutral posts as depressive.

4.2 The Experimental Results Of Multi-Feature Integration

Table 3: Results of combined features

Feature_set	Model	Accuracy	Precision	Recall	F1-Score
D+S	**RF**	**0.688**	**0.689**	**0.688**	**0.6863**
	LR	0.696	0.7196	0.696	0.6839
	KNN	0.576	0.575	0.576	0.5737
	SVM	0.536	0.5953	0.536	0.4143
D+T	RF	0.688	0.7034	0.688	0.6785
	LR	**0.752**	**0.7652**	**0.752**	**0.7471**
	KNN	0.592	0.5926	0.592	0.5858
	SVM	0.536	0.5953	0.536	0.4143
T+S	RF	0.712	0.728	0.712	0.7041
	LR	**0.768**	**0.7823**	**0.768**	**0.7634**
	KNN	0.6	0.6015	0.6	0.593
	SVM	0.536	0.5953	0.536	0.4143
D+T +S	RF	0.656	0.6564	0.656	0.6541
	LR	**0.808**	**0.8184**	**0.808**	**0.8054**
	KNN	0.592	0.592	0.592	0.5873
	SVM	0.536	0.5953	0.536	0.4143

Note: DistilBert+SWN：D+S；DistilBert+TF-IDF：D+T；TF-IDF+SWN：T+S；DistilBert+TF-IDF+SWN：D+T+S

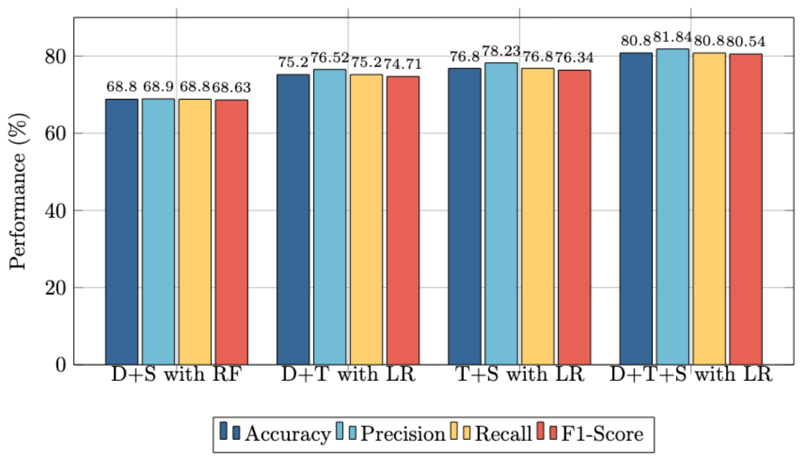

Fig. 9: The best results of multi-feature

The results of the multi-feature integration experiments(as shown in Table 3 and Fig. 9) demonstrated that combining features significantly enhances model performance, as evidenced by the superior results achieved with the integration of DistilBERT, TF-IDF, and SentiWordNet (SWN). Logistic Regression (LR) achieved an accuracy of 80.8% and an F1 score of 80.54% when leveraging all three feature sets, highlighting the efficacy of this approach. The high F1 score indicates a strong balance between precision and recall, making the proposed framework particularly suitable for real-world applications like early depression detection, where missing cases (false negatives) could have severe consequences.

The improvements observed in multi-feature integration can be attributed to the complementary strengths of the individual feature sets. DistilBERT embeddings provided semantic depth and contextual understanding, enabling the model to interpret complex linguistic expressions. TF-IDF contributed foundational lexical relevance by quantifying term importance across the dataset, while SentiWordNet captured emotional polarity, emphasizing subtle shifts in sentiment often indicative of depressive tendencies. Together, these features formed a robust multi-dimensional representation that allowed the model to effectively identify nuanced patterns in depressive text.

Pairwise combinations also demonstrated significant improvements compared to single-feature experiments. For instance, TF-IDF combined with SWN achieved an F1 score of 76.34% using LR, highlighting the effective synergy between lexical and emotional features. Similarly, DistilBERT paired with TF-IDF yielded an F1 score of 74.71%, showcasing the benefit of integrating semantic depth with lexical importance. Despite these enhancements, traditional models such as SVM and KNN continued to underperform across all feature combinations. SVM frequently misclassified neutral posts as depressive, likely because it struggles to effectively handle the complexity of high-dimensional combined feature spaces. Similarly, KNN struggled with posts exhibiting subtle emotional expressions, underscoring its reliance on proximity-based decision-making, which may fail to capture global patterns in the data.

These findings demonstrate the importance of leveraging multi-dimensional features in improving classification performance and reducing errors. For example, depressive posts with implicit emotional indicators or complex linguistic structures were better classified when all three feature sets were integrated. However, the experiments also revealed certain limitations. Posts containing strong emotional keywords but lacking depressive context were sometimes misclassified as depressive (false positives), while posts with implicit depressive indicators but neutral sentiment were occasionally misclassified as non-depressive (false negatives). These errors highlight the challenges of balancing emotional, semantic, and lexical dimensions in multi-feature integration and suggest areas for future refinement.

4.3 The Experimental Results of Feature Selection

Table 4: Results of feature selection

Feature_selection	Model	Accuracy	Precision	Recall	F1-Score
LR-RFE	RF	0.752	0.7559	0.752	0.7500
	LR	**0.800**	**0.8121**	**0.800**	**0.7969**
	KNN	0.680	0.6844	0.680	0.6757
	SVM	0.680	0.7126	0.680	0.6624

In the third experiment(the results are shown in Table 4), Logistic Regression-based Recursive Feature Elimination (LR-RFE) was applied to optimize the combined feature set of DistilBERT, TF-IDF, and SentiWordNet (SWN). Starting with a comprehensive feature set, features were iteratively removed to identify the optimal subset. The results demonstrated that reducing the feature set to 400 dimensions achieved the best balance between computational efficiency and classification accuracy. This subset effectively retained the most critical features while minimizing redundancy, demonstrating the practical utility of feature selection in handling high-dimensional datasets.

Applying LR-RFE resulted in a dimensionality reduction of approximately 75%, which significantly improved computational efficiency, reducing training time by nearly 40% while maintaining the overall performance of the model. The Logistic Regression (LR) model, for instance, maintained an accuracy of 80.0% and an F1 score of 79.69%, highlighting the practical benefits of feature selection for resource-constrained environments. Additionally, traditional models such as Random Forest (RF) and Support Vector Machines (SVM) demonstrated marked improvements. RF's F1 score increased from 65.41% to 75%, and SVM's F1 score rose from 41.43% to 66.24%, underscoring the ability of LR-RFE to streamline feature spaces and enhance generalization performance.

Beyond computational efficiency, the use of LR-RFE provided significant interpretability advantages. Logistic Regression inherently ranks feature importance through its coefficients, and LR-RFE amplifies this by isolating the most critical features. This capability is particularly valuable for mental health applications, as it allows practitioners to trace predictions back to specific linguistic, semantic, or emotional features. For example, the framework could identify whether a post's classification as depressive was driven primarily by emotional polarity (SWN), lexical patterns (TF-IDF), or semantic context (DistilBERT). Such transparency facilitates trust and practical application in real-world mental health interventions. These findings highlight LR-RFE as an essential component for deploying machine learning models in complex textual classification tasks, where scalability and explainability are paramount.

4.4 Comparative Analysis

Table 5: Comparison of Feature-based Approaches for Depression Detection

Feature_set	Model	F1-Score	Accuracy	Precision	Recall
Sentiment Features, AFINN, NRC(NRC_SA), MPQA, and SenticNet[33]	Ensemble	0.7655	0.7555	0.6550	0.7455
BOW, TF-IDF, Count Vectorization, Word2Vec[19]	SVM	0.79	0.8179	0.79	**0.82**
TF-IDF, Word embedding, N-gram[30]	SVM	0.6336	**0.9077**	0.6531	0.6153
DistilBert, SentiWordNet	RF	0.6863	0.688	0.689	0.688
DistilBert, TF-IDF	LR	0.7471	0.752	0.7652	0.752
TF-IDF, SentiWordNet	LR	0.7634	0.768	0.7823	0.768
DistilBert, SentiWordNet, TF-IDF	LR	**0.8054**	0.808	**0.8184**	0.808

The results from the three experiments collectively highlight the critical role of feature extraction, integration, and selection techniques in addressing the challenges of depression detection. Each experiment emphasizes a critical aspect of the framework's performance, providing deeper insights into how different methods contribute to better classification results. A comparison with previous studies (as shown in Table 5) shows that this approach outperforms other feature combinations, highlighting the importance of feature engineering in model performance. Specifically:

The single-feature experiments underscored the strengths and limitations of individual feature types. DistilBERT embeddings emerged as the most effective single feature, with Logistic Regression (LR) achieving an F1 score of 74.4%, emphasizing the utility of deep semantic representations in capturing contextual nuances. Similarly, SentiWordNet (SWN) demonstrated the value of emotional polarity features, particularly when paired with Random Forest (RF), achieving an F1 score of 67.86%. However, models like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) consistently underperformed due to their limited ability to handle high-dimensional feature spaces, revealing a gap that necessitates feature optimization and model refinement.

The multi-feature integration experiments further validated the complementary nature of combining lexical, semantic, and emotional features. The integration of DistilBERT, TF-IDF, and SWN resulted in the highest performance, with LR achieving an F1 score of 80.54%. This demonstrates that combining multi-dimensional features enables the model to capture diverse aspects of depressive text, from semantic depth to emotional polarity and lexical relevance. The results also show that multi-feature integration reduces classification errors, particularly in ambiguous cases, by allowing the model to account for relationships across different layers of information.

The feature selection experiment using Logistic Regression-based Recursive Feature Elimination (LR-RFE) highlighted the impact of optimizing high-dimensional feature sets. Reducing the feature set to 400 dimensions improved computational efficiency by approximately 40% while maintaining robust classification performance. LR-RFE proved particularly effective in enhancing the adaptability of traditional models, with RF's F1 score increasing from 65.41% to 75%, and SVM's F1 score improving from 41.43% to 66.24%. Additionally, the ability of LR-RFE to rank feature importance contributes to the interpretability of the framework, which is especially beneficial for understanding the underlying factors driving model predictions.

In summary, the experiments collectively highlight the importance of combining advanced feature extraction techniques, multi-dimensional feature integration, and feature selection optimization. Additionally, as this study did not perform parameter optimization, the experimental results reflect the raw effectiveness of the feature sets and models, leaving potential room for future hyperparameter tuning and further research..

5.0 CONCLUSION

This study investigated the effectiveness of multi-dimensional feature extraction and feature selection techniques for depression detection using social media text. By integrating TF-IDF with N-grams, DistilBERT embeddings, and SentiWordNet, the proposed framework successfully captured linguistic, semantic, and emotional features. The application of logistic regression-based recursive feature elimination (LR-RFE) further optimized the high-dimensional feature set, enhancing computational efficiency and improving model performance. The findings highlight three key insights. First, individual feature extraction methods showed varied effectiveness. Deep semantic features, such as DistilBERT embeddings, excelled in capturing contextual nuances, while emotion polarity features from SentiWordNet provided valuable insights into emotional expressions. However, traditional machine learning models like KNN and SVM struggled with high-dimensional feature spaces, underperforming relative to Logistic Regression (LR) and Random Forest (RF). Second, multi-feature integration proved highly effective, with the combined feature set outperforming any single feature set. The complementary strengths of diverse features enabled the model to capture complex patterns in depressive text, reducing classification errors and improving robustness. This underscores the importance of leveraging multi-dimensional features to address the intricate nature of depressive language. Finally, feature selection, particularly LR-RFE, played a critical role in refining the combined feature set. It reduced redundancy, identified the most informative features, and significantly improved the interpretability and generalization of traditional models, such as RF and SVM.

Despite these contributions, this study has limitations. The reliance on a single dataset restricts the generalizability of the findings, and the use of default model parameters leaves room for optimization. Additionally, the framework was not evaluated across multiple platforms or with advanced domain-specific embeddings, limiting its applicability to broader contexts. Future research should address these limitations by validating the framework across diverse datasets and social media platforms, applying hyperparameter tuning, and exploring the integration of advanced deep learning models with domain-specific embeddings. These efforts could further enhance classification performance and scalability, paving the way for more robust applications in mental health analytics.

ACKNOWLEDGMENT

The authors acknowledge the Geran Galakan Penyelidik Muda (GGPM), grant number GGPM-2022-057 funded by the Universiti Kebangsaan Malaysia.

REFERENCES

[1] K. M. Hasib, M. R. Islam, S. Sakib, Md. A. Akbar, I. Razzak, and M. S. Alam, ‘Depression Detection From Social Networks Data Based on Machine Learning and Deep Learning Techniques: An Interrogative Survey’, *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1568–1586, Aug. 2023, doi: 10.1109/TCSS.2023.3263128.

[2] J. Zhong, W. Du, L. Zhang, H. Peng, and B. Hu, ‘Feature extraction based on sparse graphs embedding for automatic depression detection’, *Biomedical Signal Processing and Control*, vol. 86, p. 105257, Sep. 2023, doi: 10.1016/j.bspc.2023.105257.

[3] D. William and D. Suhartono, ‘Text-based Depression Detection on Social Media Posts: A Systematic Literature Review’, *Procedia Computer Science*, vol. 179, pp. 582–589, Jan. 2021, doi: 10.1016/j.procs.2021.01.043.

[4] K. Feng *et al.*, ‘Effects of music therapy on major depressive disorder: A study of prefrontal hemodynamic functions using fNIRS’, *Psychiatry Research*, vol. 275, pp. 86–93, May 2019, doi: 10.1016/j.psychres.2019.03.015.

[5] S. Cunningham, C. C. Hudson, and K. Harkness, ‘Social Media and Depression Symptoms: a Meta-Analysis’, *Res Child Adolesc Psychopathol*, vol. 49, no. 2, pp. 241–253, Feb. 2021, doi: 10.1007/s10802-020-00715-7.

[6] F. Rice *et al.*, ‘Adolescent and adult differences in major depression symptom profiles’, *Journal of Affective Disorders*, vol. 243, pp. 175–181, Jan. 2019, doi: 10.1016/j.jad.2018.09.015.

[7] M. D. Aron Halfin, ‘Depression: The Benefits of Early and Appropriate Treatment’, vol. 13, Nov. 2007, Accessed: Dec. 12, 2024. [Online]. Available: https://www.ajmc.com/view/nov07-2638ps092-s097

[8] F. Crestani, D. E. Losada, and J. Parapar, *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project*. Springer Nature, 2022.

[9] M. Squires *et al.*, ‘Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment’, *Brain Inf.*, vol. 10, no. 1, p. 10, Apr. 2023, doi: 10.1186/s40708-023-00188-6.

[10] A. Ahmed *et al.*, ‘Machine learning models to detect anxiety and depression through social media: A scoping review’, *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100066, Jan. 2022, doi: 10.1016/j.cmpbup.2022.100066.

[11] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gómez, ‘A text classification framework for simple and effective early depression detection over social media streams’, *Expert Systems with Applications*, vol. 133, pp. 182–197, Nov. 2019, doi: 10.1016/j.eswa.2019.05.023.

[12] K. Schneider *et al.*, ‘Syntactic complexity and diversity of spontaneous speech production in schizophrenia spectrum and major depressive disorders’, *Schizophr*, vol. 9, no. 1, Art. no. 1, May 2023, doi: 10.1038/s41537-023-00359-8.

[13] N. Vedula and S. Parthasarathy, ‘Emotional and linguistic cues of depression from social media’, in *Proceedings of the 2017 International Conference on Digital Health*, 2017, pp. 127–136.

[14] M.-H. Le-Nguyen, F. Turgis, P.-E. Fayemi, and A. Bifet, ‘Real-time learning for real-time data: online machine learning for predictive maintenance of railway systems’, *Transp. Res. Procedia*, vol. 72, pp. 171–178, Jan. 2023, doi: 10.1016/j.trpro.2023.11.391.

[15] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, ‘A textual-based featuring approach for depression detection using machine learning classifiers and social media texts’, *Computers in Biology and Medicine*, vol. 135, p. 104499, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104499.

[16] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, ‘Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression’, in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2017, pp. 138–140. doi: 10.1109/ICTC.2017.8190959.

[17] E. Campillo-Ageitos, J. Martinez-Romo, and L. Araujo, ‘UNED-MED at eRisk 2022: depression detection with TF-IDF, linguistic features and Embeddings’.

[18] J. Liu and M. Shi, ‘A Hybrid Feature Selection and Ensemble Approach to Identify Depressed Users in Online Social Media’, *Frontiers in Psychology*, vol. 12, 2022, Accessed: Jun. 25, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.802821

[19] D. Muzafar, F. Y. Khan, and M. Qayoom, ‘Machine Learning Algorithms for Depression Detection and Their Comparison’, Jan. 09, 2023, *arXiv*: arXiv:2301.03222. doi: 10.48550/arXiv.2301.03222.

[20] F. Noor, M. Iqbal, S. Mehmood, A. Jaffar, S. Yaqoob, and I. Ul Haq, ‘DEPRESSION DETECTION IN SOCIAL MEDIA USING BAGGING CLASSIFIER’, *Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)*, vol. 42, pp. 53–75, Jan. 2023, doi: 10.17605/OSF.IO/EVXB7.

[21] J. C. Eichstaedt *et al.*, ‘Facebook language predicts depression in medical records’, *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11203–11208, Oct. 2018, doi: 10.1073/pnas.1802331115.

[22] J. de Godoi Brandão and W. P. Calixto, ‘N-Gram and TF-IDF for Feature Extraction on Opinion Mining of Tweets with SVM Classifier’, in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Sep. 2019, pp. 1–5. doi: 10.1109/IDAP.2019.8875900.

[23] V. Dogra, A. Singh, S. Verma, Kavita, N. Z. Jhanjhi, and M. N. Talib, ‘Analyzing DistilBERT for Sentiment Classification of Banking Financial News’, in *Intelligent Computing and Innovation on Data Science*, S.-L. Peng, S.-Y. Hsieh, S. Gopalakrishnan, and B. Duraisamy, Eds., in 网络和系统系列丛书中的讲义. Singapore: Springer Nature, 2021, pp. 501–510. doi: 10.1007/978-981-16-3153-5_53.

[24] M. Husnain, M. M. S. Missen, N. Akhtar, M. Coustaty, S. Mumtaz, and V. B. S. Prasath, ‘A systematic study on the role of SentiWordNet in opinion mining’, *Front. Comput. Sci.*, vol. 15, no. 4, p. 154614, Jun. 2021, doi: 10.1007/s11704-019-9094-0.

[25] P. Kumar, P. Samanta, S. Dutta, M. Chatterjee, and D. Sarkar, ‘Feature Based Depression Detection from Twitter Data Using Machine Learning Techniques’, *JSR*, vol. 66, no. 02, pp. 220–228, 2022, doi: 10.37398/JSR.2022.660229.

[26] S. Samanvitha, A. R. Bindiya, S. Sudhanva, and B. S. Mahanand, ‘Naïve Bayes Classifier for depression detection using text data’, in *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*, Mysuru, India: IEEE, Dec. 2021, pp. 418–421. doi: 10.1109/ICEECCOT52851.2021.9708014.

[27] J. Sharma and V. Tomer, ‘Depression detection using sentiment analysis of social media data’, presented at the INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE “TECHNOLOGY IN AGRICULTURE, ENERGY AND ECOLOGY” (TAEE2022), Dushanbe, Republic of Tajikistan, 2022, p. 020044. doi: 10.1063/5.0104192.

[28] H. Burkhardt, M. Pullmann, T. Hull, P. Areán, and T. Cohen, ‘Comparing emotion feature extraction approaches for predicting depression and anxiety’, in *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, A. Zirikly, D. Atzil-Slonim, M. Liakata, S. Bedrick, B. Desmet, M. Ireland, A. Lee, S. MacAvaney, M. Purver, R. Resnik, and A. Yates, Eds., Seattle, USA: Association for Computational Linguistics, Jul. 2022, pp. 105–115. doi: 10.18653/v1/2022.clpsych-1.9.

[29] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, ‘Detection of Depression-Related Posts in Reddit Social Media Forum’, *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.

[30] M. Stankevich, V. Isakov, D. Devyatkin, and I. Smirnov, ‘Feature Engineering for Depression Detection in Social Media’:, in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 426–431. doi: 10.5220/0006598604260431.

[31] S. H. Hosseini-Saravani, S. Besharati, H. Calvo, and A. Gelbukh, ‘Depression Detection in Social Media Using a Psychoanalytical Technique for Feature Extraction and a Cognitive Based Classifier’, in *Advances in Computational Intelligence*, L. Martínez-Villaseñor, O. Herrera-Alcántara, H. Ponce, and F. A. Castro-Espinoza, Eds., in 计算机科学系列书籍讲义. Cham: Springer International Publishing, 2020, pp. 282–292. doi: 10.1007/978-3-030-60887-3_25.

[32] R. Chiong, G. S. Budhi, and S. Dhakal, ‘Combining Sentiment Lexicons and Content-Based Features for Depression Detection’, *IEEE Intell. Syst.*, vol. 36, no. 6, pp. 99–105, Nov. 2021, doi: 10.1109/MIS.2021.3093660.

[33] L. Ansari, S. Ji, Q. Chen, and E. Cambria, ‘Ensemble Hybrid Learning Methods for Automated Depression Detection’, *IEEE Transactions on Computational Social Systems*, vol. 10, no. 1, pp. 211–219, Feb. 2023, doi: 10.1109/TCSS.2022.3154442.

[34] J. Ye *et al.*, ‘Analysis and Recognition of Voluntary Facial Expression Mimicry Based on Depressed Patients’, *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 8, pp. 3698–3709, Aug. 2023, doi: 10.1109/JBHI.2023.3260816.

[35] C. Fang, G. Dianatobing, T. Atara, I. S. Edbert, and D. Suhartono, 'Feature Extraction Methods for Depression Detection Through Social Media Text', in *2022 6th International Conference on Informatics and Computational Sciences (ICICoS)*, Sep. 2022, pp. 117–121. doi: 10.1109/ICICoS56336.2022.9930596.

[36] L. Ansari, S. Ji, Q. Chen, and E. Cambria, 'Ensemble Hybrid Learning Methods for Automated Depression Detection', *IEEE Transactions on Computational Social Systems*, vol. 10, no. 1, pp. 211–219, Feb. 2023, doi: 10.1109/TCSS.2022.3154442.

[37] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, 'A textual-based featuring approach for depression detection using machine learning classifiers and social media texts', *Computers in Biology and Medicine*, vol. 135, p. 104499, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104499.

[38] M. Janatdoust, F. Ehsani-Besheli, and H. Zeinali, 'KADO@LT-EDI-ACL2022: BERT-based Ensembles for Detecting Signs of Depression from Social Media Text', in *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 265–269. doi: 10.18653/v1/2022.ltedi-1.38.

[39] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, 'Detection of Depression-Related Posts in Reddit Social Media Forum', *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.

[40] D. E. Losada, F. Crestani, and J. Parapar, 'eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations', in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, G. J. F. Jones, S. Lawless, J. Gonzalo, L. Kelly, L. Goeuriot, T. Mandl, L. Cappellato, and N. Ferro, Eds., in 计算机科学系列讲义. Cham: Springer International Publishing, 2017, pp. 346–360. doi: 10.1007/978-3-319-65813-1_30.

[41] D. E. Losada, F. Crestani, and J. Parapar, 'Overview of eRisk: Early Risk Prediction on the Internet', in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, P. Bellot, C. Trabelsi, J. Mothe, F. Murtagh, J. Y. Nie, L. Soulier, E. SanJuan, L. Cappellato, and N. Ferro, Eds., in 计算机科学系列讲义. Cham: Springer International Publishing, 2018, pp. 343–361. doi: 10.1007/978-3-319-98932-7_30.

[42] D. E. Losada and F. Crestani, 'A Test Collection for Research on Depression and Language Use', in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, vol. 9822, N. Fuhr, P. Quaresma, T. Gonçalves, B. Larsen, K. Balog, C. Macdonald, L. Cappellato, and N. Ferro, Eds., in Lecture Notes in Computer Science, vol. 9822. , Cham: Springer International Publishing, 2016, pp. 28–39. doi: 10.1007/978-3-319-44564-9_3.

[43] S. Elbagir and J. Yang, 'Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment', *Hong Kong*, 2019.

[44] F. Illia, M. P. Eugenia, and S. A. Rutba, 'Sentiment Analysis on PeduliLindungi Application Using TextBlob and VADER Library', *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2021, no. 1, Art. no. 1, 2021, doi: 10.34123/icdsos.v2021i1.236.

[45] V. Bonta, N. Kumaresh, and J. Naulegari, 'A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis', *Asian Journal of Computer Science and Technology*, vol. 8, pp. 1–6, Mar. 2019, doi: 10.51983/ajcst-2019.8.S2.2037.

[46] D. Cortiz, 'Exploring Transformers models for Emotion Recognition: a comparision of BERT, DistilBERT, RoBERTa, XLNET and ELECTRA', in *Proceedings of the 2022 3rd International Conference on Control, Robotics and Intelligent System*, in CCRIS '22. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 230–234. doi: 10.1145/3562007.3562051.

[47] T. Zhang, K. Yang, S. Ji, and S. Ananiadou, 'Emotion fusion for mental illness detection from social media: A survey', *Information Fusion*, vol. 92, pp. 231–246, Apr. 2023, doi: 10.1016/j.inffus.2022.11.031.

[48] C. Dev and A. Ganguly, ‘Sentiment Analysis of Assamese Text Reviews: Supervised Machine Learning Approach with Combined n-gram and TF-IDF Feature’, *ADBU Journal of Electrical and Electronics Engineering (AJEEE)*, vol. 5, no. 2, pp. 18–30, Sep. 2023.

[49] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, ‘DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter’, arXiv.org. Accessed: Dec. 05, 2023. [Online]. Available: https://arxiv.org/abs/1910.01108v4

[50] S. Bhadra and C. J. Kumar, ‘Enhancing the efficacy of depression detection system using optimal feature selection from EHR’, *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 27, no. 2, pp. 222–236, 2023, doi: 10.1080/10255842.2023.2181660.

[51] Md. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, ‘Depression detection from social network data using machine learning techniques’, *Health Inf Sci Syst*, vol. 6, no. 1, p. 8, Aug. 2018, doi: 10.1007/s13755-018-0046-0.

[52] H. Yu and S. Kim, ‘SVM Tutorial: Classification, Regression, and Ranking’, *Handbook of Natural Computing*, Jan. 2012, doi: 10.1007/978-3-540-92910-9_15.

[53] P. Mishra and G. U. Devi, ‘Performance investigation for analysis of Predicting and understanding human depression behavior from social network analysis using SVM model: A Review’, in *2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, Feb. 2023, pp. 1–6. doi: 10.1109/IEMECON56962.2023.10092317.

[54] L. Breiman, ‘Random Forests’, *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[55] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, ‘Early Detection of Depression: Social Network Analysis and Random Forest Techniques’, *Journal of Medical Internet Research*, vol. 21, no. 6, p. e12554, Jun. 2019, doi: 10.2196/12554.

[56] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, ‘KNN Model-Based Approach in Classification’, in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, R. Meersman, Z. Tari, and D. C. Schmidt, Eds., in 计算机科学系列书籍讲义. Berlin, Heidelberg: Springer, 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[57] N. B. Oğur *et al.*, ‘Detection of depression and anxiety in the perinatal period using Marine Predators Algorithm and kNN’, *Computers in Biology and Medicine*, vol. 161, p. 107003, Jul. 2023, doi: 10.1016/j.compbiomed.2023.107003.

[58] E. Bisong, ‘Logistic Regression’, in *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, E. Bisong, Ed., Berkeley, CA: Apress, 2019, pp. 243–250. doi: 10.1007/978-1-4842-4470-8_20.

[59] Z. Guo, N. Ding, M. Zhai, Z. Zhang, and Z. Li, ‘Leveraging Domain Knowledge to Improve Depression Detection on Chinese Social Media’, *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1528–1536, Aug. 2023, doi: 10.1109/TCSS.2023.3267183.

[60] D. M. W. Powers, ‘Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation’, Oct. 10, 2020, *arXiv*: arXiv:2010.16061. doi: 10.48550/arXiv.2010.16061.

[61] A. Tharwat, ‘Classification assessment methods’, *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003.