National College of Ireland

A Hybrid Approach to NLP-Based Depression Detection: Integrating BERT and Word2Vec

MSc Research Project
Data Analytics

Yoshitha Ganji
Student ID: 23102268

School of Computing
National College of Ireland

Supervisor:Abubakr Siddig

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……Yoshitha Ganji……

Student ID: ……23102268……

Programme: ……Ms Data Analytics…… **Year:** 2023-2024……

Module: ……Researchpaper……

Supervisor: …Abubhakar Siddig……

Submission Due Date: ……12-08-2024……

Project Title: A Hybrid Approach to NLP-Based Depression Detection: Integrating BERT and Word2Vec ……

Word Count: ……7813…… **Page Count**…21……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Yoshitha Ganji……

Date: ……10/8/2024……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Hybrid Approach to NLP-Based Depression Detection: Integrating BERT and Word2Vec

Yoshitha Ganji
23102268

Abstract

This study takes an innovative approach to detecting depression and suicidal ideation through social media by integrating BERT and Word2Vec into a powerful hybrid model. This model, augmented by analyses of CNN, LSTM, BiLSTM, and RoBERTa architectures, aims to substantially elevate accuracy in identifying the nuances of depressive expressions often overlooked by traditional single-model methods. Motivated by the vast potential of advanced NLP technologies, our comprehensive strategy harnesses BERT's exceptional contextual understanding and Word2Vec's in-depth semantic analysis. This dual approach addresses the complex challenge of decoding the myriad emotional cues within social media text that signal mental health concerns. This research journey, ambitious in its scope, not only pioneered the development of the hybrid model but also critically assessed its performance against othersignificant neural networks. This broad analysis confirmed the hybrid model's superior capabilityin capturing the subtleties of emotional states online, achieving a remarkable 0.90 accuracy rate. This work contributes significantly to both theoretical and practical domains. Theoretically, it opens new dialogues about the synergy between various NLP models for more accurate emotional detection, establishing new standards in the field. Practically, it signals a breakthroughin creating real-time monitoring tools for early mental health intervention on social media, offering scalable solutions that could revolutionize public health strategies.

1: Introduction

1.1 Introduction to the Research Problem

Affecting tens of millions of people worldwide, depression is an alternative successful clinical situation. It stands proud for a constant feeling of gloom, a lack of hobbies, and hassle with everyday responsibilities. According to the World Health Organization (WHO), approximately 264 million humans globally experience hopelessness; intense instances may result in suicide. Since social media has created new channels for the expression of thoughts and emotions, it is a helpful device for identifying depressive signs and symptoms. But as Beniwal and Saraswat (2024) are aware, the exceptional variety of minor changes in human language makes it tough to spot loss in social media messages. With practical uses ranging from sentiment evaluation to highbrow health evaluation, machine learning (ML) and natural language processing (NLP) are evolving as aspirational text assessment gear. Although those buildings have super capacity, it is

far uncertain how they may hit upon despair in social media posts (Singh *et al.*, 2023). This work analyses the ability of two famous herbal language processing (NLP) models, Word2Vec and BERT (Bidirectional Encoder Representations from Transformers), to discover sadness in social media analysis so that it can close this gap. Using an evaluation of many models, this study aims to discover which one performs better and the reasons behind it, therefore helping with the introduction of stronger contraptions for mental health monitoring and intervention.

1.2 Significance of the Research

Examining how depression is expressed on social media is critical for several unique reasons. It improves the potential for proper tracking and dealing with intellectual health issues. When despair symptoms are consistently identified on social media, those in need can get brief and efficient assistance, consequently relieving some stress on healthcare systems and improving affected person results. This work makes use of machine learning (ML) and natural language processing (NLP) to help create tools for early diagnosis and intervention, thereby supplying an initiative-taking approach to mental fitness remedy (NA and Veni, 2022). Second, our workhelps to develop NLP and machine learning commonly by highlighting the benefits and downsides of BERT and Word2Vec in a certain software situation. Understanding how nicely these algorithms perceive sorrow will lead to similar studies and sensible uses in different NLP fields. Moreover, by stressing the subtleties of these models, we help to build stronger and more effective NLP-based intellectual health detection systems (Rizwan *et al*. 2022).

1.3 Research Question

The essential question driving this work is, "To what quantity can a comparative evaluation among BERT and Word2Vec algorithms, as implemented for the detection of melancholy using social media posts, display their respective strengths and weaknesses?" Moreover, how can the understanding acquired from this evaluation be conducted to design a better or hybrid version that makes most use of the benefits of numerous algorithms in depression prognosis?" By approaching an evaluation of the efficacy of BERT and Word2Vec in identifying depressive symptoms using social media facts, this observer goals to check whether the version performs better and what generates performance versions. As cited by Zhang *et al*. (2022), it also looks at developing a hybrid model combining the blessings of both strategies to permit more unique and dependable melancholy detection. This work aims to increase the domain of NLP-based total mental health detection and offer an intensive understanding of these methods.

1.4 Proposed Solution

Examining the capabilities of BERT and Word2Vec to become aware of depression in social media postings gives a thorough solution to the research question. This comparison is achieved using a delegated series of exams aimed at assessing the models' memory, accuracy, precision, and model efficacy in recognizing despair signs, symptoms, and indicators. Examining these criteria helps one to evaluate how successfully the version has evolved in spotting fundamental

causes of conduct and miserable symptoms (Arora et al., 2024). The intention could be to create a hybrid variation combining BERT and Word2Vec depending on a comparative analysis. This hybrid technique seeks to combine bidirectional contextual awareness from BERT with low-fee, actual vector representations from Word2Vec. Integrating those models, consistent with Sharma *et al*. (2023), seeks to increase the accuracy and dependability of despair detection, therefore generating a more useful tool for tracking and treating intellectual health issues. This answer seeks to make contributions to the improvement of green NLP-based answers for initiative- taking intellectual fitness as well as a useful resource and early identification with the aid of filling in a chief want in present methods.

1.5 Expected Contributions

Natural language processing (NLP) and intellectual health detection must be a great deal superior using this work. First and most importantly, they may cross over the blessings and disadvantages of each variant by way of intently reading Word2Vec and BERT inside the framework of social media melancholy detection. These effects will help professionals choose the right approach for their unique necessities. Second, it is widely thought that developing a hybrid model combining the excellent functions of Word2Vec and BERT will improve despair detection consistency and accuracy (Chaturvedi, 2023). This improvement ought to lead to better equipment for monitoring intellectual health, consequently allowing early identification and treatment of conditions.. The predominant purpose of this work is to raise NLP's possibilities for usage in mental fitness packages, therefore allowing its main development in early despair diagnosis and treatment.

1.6 Structure of the Document

The office work is set out as follows: The research topic, its importance, the recommended answer, and the predicted contributions are compiled in Chapter 1, Introduction. Reviewing the literature in Chapter 2 highlighting research gaps and despair depression detection employing BERT and Word2Vec, Chapter 3, Methodology, is going into an exquisite period at the techniques applied to create the hybrid model and contrast Word2Vec with BERT. Chapter 4: Results and Discussion suggests and analyses the experimental information collectively with a comparison of the models. A summary of the work, an evaluation of the outcomes, and suggestions for further reading are provided in Chapter 5, Conclusion and Further Work.

2: Literature Review

2.1 Introduction to Literature Review

Since it offers a thorough assessment of earlier studies on the usage of Natural Language Processing (NLP) models, appreciably Word2Vec and BERT, for melancholy diagnosis, the literature evaluation is a crucial part of this research. The occurrence and results of depression, the use of social media in intellectual health monitoring, and the usage of herbal language processing (NLP) strategies for social media textual content analysis for mental health signs are compiled in this chapter from earlier studies. By way of a comparison of the blessings and

downsides of BERT and Word2Vec in this context, this literature evaluation aims to focus on the possibility of creating a hybrid model and, as a result, identify holes in gift methodologies. The chapter is constructed systematically to provide a methodical basis for comprehending the study issue, therefore supporting the importance of this look at and guiding the next methodological technique. This thorough study offers the shape for looking at sparkling ideas to use the contemporary NLP model to diagnose depression.

2.2 Depression and Mental Health

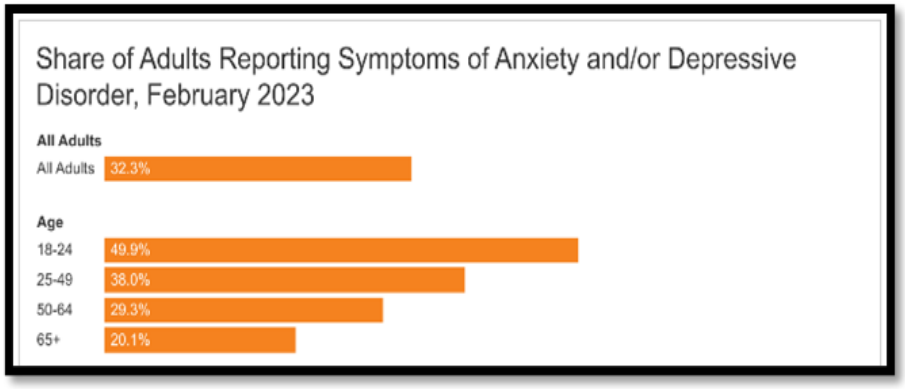

Figure 1: Statistics of Anxiety and depressive disorder globally
(Source: Ellouze and Hadrich Belguith, 2022)

Depression is a critical clinical condition characterized by a sustained low mood, alteration in sleep pattern, appetite fluctuations, and loss of interest in activities that have been previously enjoyed. The symptoms cause significant impairment in daily functioning continuously over time. According to Guo et al. 2024, the change in appetite leads to an unhealthy weight change. Social interactions become sources of anxiety and generally increase feelings of loneliness and isolation. As reported by the World Health Organization, globally, more than 264 million people of all ages suffer from depression, which is also the leading cause of disability worldwide. According to Ellouze and Hadrich Belguith, discussion 2022, women are more affected than men. Economically, the cost of depression is huge in healthcare expenditure and lost productivity. Figure 1 illustrates the global statistics of anxiety and depressive disorders.

2.3 Social Media and Mental Health

Social media platforms have become integral spaces for real-time expression and interaction, fostering both positive and negative impacts on mental health. While users can share experiences and seek support online, excessive use has been associated with heightened feelings of depression (Harati et al., 2022). The impact of social media on mental health, particularly among youth, is significant (Figure 2).The constant exposure to idealized images and social comparisons prevalent on these platforms can significantly impact users' self-esteem.

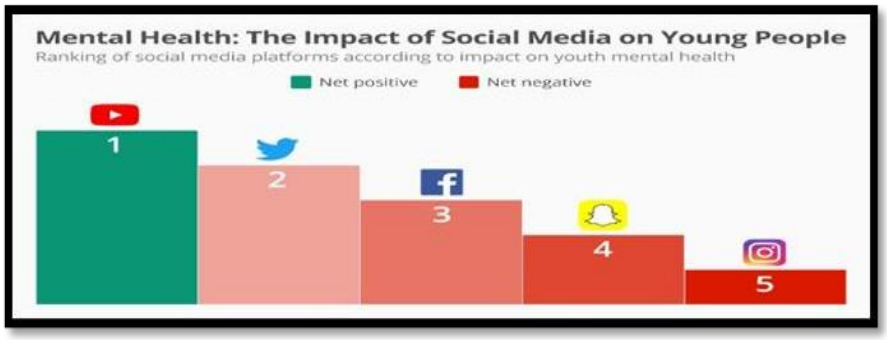

Figure 2: Impact of social media on mental health
(Source: Harati et al. 2022)

Despite the risks, social media data holds promise for mental health research, as indicated by Arowosegbe and Oyelade (2023), who emphasize its potential for identifying linguistic markers and emotional cues related to depression, paving the way for early intervention. However, challenges persist in accurately interpreting emotional content due to informal language and slang commonly used in online communication, as highlighted by MohammadiBaghmolaei and Ahmadi (2023), who also note the discrepancies between online personas and actual mental health struggles. Despite these obstacles, ongoing research in this area offers the potential for valuable tools in monitoring mental health and detecting depression through social media data analysis.

2.4 Natural Language Processing (NLP) in Mental Health

The conflict against intellectual infection, natural language processing NLP has turned out to be a quite powerful weapon. As a subset of artificial intelligence (AI) and natural language processing (NLP), computers could interpret and modify human language, therefore offering a unique perspective for inspecting the sizeable textual records produced on the internet (Jain et al.2024). This statistic has unrealized ability, particularly from social media sites, to show traits andsigns connected with mental health issues like depression.

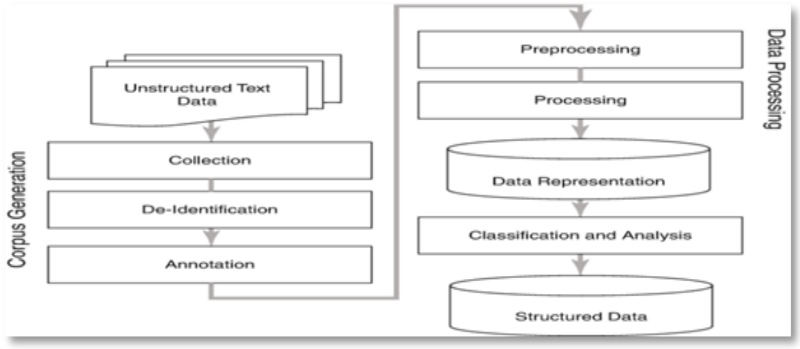

Figure 3: NLP in mental health (Source: Jothi and Pandeeswari, 2023)

Natural Language Processing (NLP) tools enable researchers to analyze social media language nuances. Sentiment analysis, a basic NLP technique, categorizes text as positive, negative, or neutral. Jothi and Pandeeswari (2023) note this helps identify posts expressing negative

emotions. Natural Language Processing (NLP) offers a range of tools for analyzing social media language in the context of mental health (Figure 3)."Topic modelling uncovers underlying themes in social media content.Text classification, trained on large datasets, detects depression-related patterns in posts. However, NLP faces challenges with informal language, sarcasm, and acronyms on social media. Yang (2023) points out that users' curated onlinepersonas complicate distinguishing genuine depression admissions from performative sharing.Despite limitations, NLP remains valuable for mental health condition detection in socialmedia data. As NLP techniques evolve to handle complex human language, they become increasingly important in mental health research and treatment, potentially improving earlydiagnosis and intervention outcomes**.**

2.5 BERT in Depression Detection

Thanks to its remarkable contextual awareness, BERT excels in NLP-based depression diagnosis. Several studies have underscored its efficacy in perceiving individuals who may be depressed and understanding social media interactions. For instance, a study using BERT to detect linguistic nuances indicative of depression in social media posts outperformed traditional NLP models, demonstrating considerable accuracy (Rani and Jain, 2024).

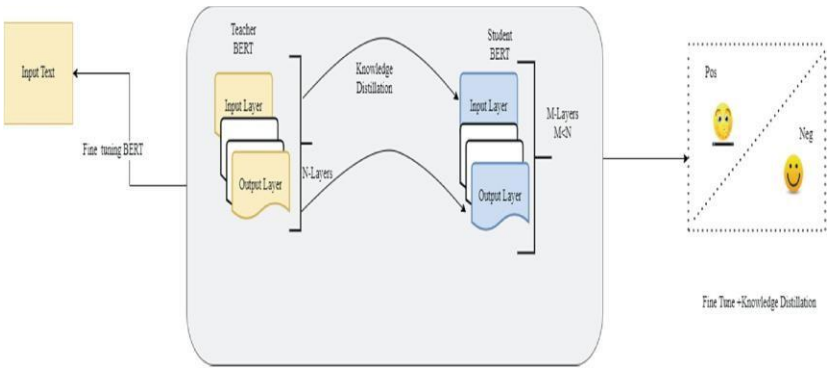

Figure 4: Usage of BERT model for depression detection
(Source: Sabaneh et al. 2023)

However, despite its advantages, the analysis of depression with BERT is not without challenges. It requires large volumes of labeled data to fully comprehend the subtleties of human language, raising concerns regarding the lack of systematic diagnosis of depression in social media data (Abuol et al. 2024). Additionally, BERT may encounter difficulty in recognizing informal language, including acronyms and slang, potentially leading to misinterpretation of emotional tone in online conversations.

Looking Ahead:

Research continues to enhance BERT's depression detection capabilities through methods such as transfer learning and addressing bias reduction in training data, as investigated by Sabaneh et al. (2023).BERT's application in depression detection is illustrated in Figure 4, showcasing its ability to process and analyze text for indicators of depression.". Despite its drawbacks, ongoing research demonstrates BERT's potential as a valuable tool for early intervention in mental health.

2.6 Word2Vec in Depression Detection

In the realm of identifying melancholy, Word2Vec, while easier to use, has garnered substantial recognition as a potent NLP model. It focuses on uncovering the underlying themes and vocabulary associated with somber language in social media posts by emphasizing semantic linkages between words. Previous studies have explored Word2Vec's potential to detect depression. For instance, Benítez-Andrades et al. (2023) conducted an examination where Word2Vec was instrumental in identifying phrase clusters frequently occurring in depression support groups, shedding light on pairings related to depressive signs.

2.7 Research Gaps and Future Direction

The use of NLP for depression diagnosis reveals significant gaps. A major challenge is the lack of labelled data, introducing bias during model training due to the absence of clinical diagnoses in social media data. Additionally, informal language, including sarcasm and slang, complicates the accuracy of current NLP models (Pandey & Vishwakarma, 2023). Overcoming these limitations will require establishing standardized labelling and collaborating with mental health professionals to obtain specific diagnosis-based datasets. Furthermore, the integration of hybrid models such as Word2Vec and BERT holds promise for more comprehensive detection strategies. For example, a hybrid model may also use Word2Vec to perceive the frequency and semantic correlations of melancholy-associated phrases and BERT's contextual recognition to assess a put-up's emotional tone (Alomari, 2024). Combining those two procedures would cause extra state-of-the-art expertise on ways depressed human beings use language on social media, consequently enhancing the effectiveness of early identity programs. Future studies must, as a result, focus on overcoming statistical obstacles, addressing informal language, and investigatinghybrid models. By bridging these gaps and enhancing the field of social media analysis for despair analysis, NLP enables early intervention and better mental health consequences for people who are suffering in silence.

2.8 Summary

This section covered a radical evaluation of research on natural language processing techniques' use in identifying despair in social media records. The study checked out the numerous factors of despair as well as how they affect intellectual well-being. The study then went over howlanguage evaluation may be used to spot social networking website despair signs. NLP emerged as a powerful tool for this purpose, with BERT and Word2Vec as leading models. WhileWord2Vec shines in recognizing subject matters, BERT is an exact substitute for compiling backdrops. Comparative critiques underlined their several benefits as well as the possibility for hybrid models in the future. Research constraints incorporate problems with categorized data regulations and casual language. Investigating hybrid models and addressing these gaps may be the main priorities of similar studies. This work is crucial as it advances the introduction of more exact and dependable NLP-based systems for social media platform despair detection.Below Table 1 shows the summary of key findings.

Author Name	Year	Key -Findings	Methodology
Beniwal and Saraswat	2024	Hybrid BERT-CNN approach improved depression detection accuracy	Combined BERT embeddings with CNN for multimodal data analysis
Alomari	2024	Hybrid models for comprehensive detection strategies	Exploration of combining Word2Vec and BERT for enhanced depression detection
Rani and Jain 2024	2024	BERT for depression detection	Evaluation of BERT's effectiveness in detecting linguistic nuances in depression
Arowosegbe and Oyelade	2023	Linguistic markers for depression detection	Utilization of social media data for early detection of depression markers

Table 1: Summary table of key studies in literature review.

3.Methodology

In the Methodology chapter, the research under scrutiny includes design, data collection, analysis, and execution phases. Addressing crucial scientific inquiry aspects, it emphasizesmethodology as the cornerstone for ensuring the research's systematic, replicable, and foundational validity. The elaboration within this section illustrates how the adopted methods align with and support the study's objectives and queries, hence advising on methodologies conducive to achieving dependable and authentic findings, which, in essence, elevate the research's overall integrity.

3.1 Research Design

In this research, Kumar et al. (2023) employs a comprehensive methodological approach, incorporating both qualitative and quantitative methods to thoroughly explore and address the research problem, objectives, and queries. This dual-faceted design facilitates the collection and analysis of vast, varied information sources, ensuring a broad examination of the phenomenon under study.

Figure 5: Types of research designs
(Source: Chakraborty and Das, 2023)

Qualitatively, it focuses on contextual sensitivity, allowing for the identification of recurring themes and patterns in textual data. Quantitatively, it harnesses the analytical power of BERT and Word2Vec algorithms for in-depth sentiment analysis and topic modelling.The core components of a research design, as illustrated in Figure 5, encompass various interconnected elements requiring careful consideration (Chakraborty and Das, 2023).

According to Chakraborty and Das (2023), this mixed-method strategy is crucial for achieving both extensive and precise insight, balancing the need for wide-ranging data encompassment with the detailed, qualitative understanding of social media dynamics. Such a methodology notably enhances research validity by mitigating bias, fostering a robust theoretical and practical understanding of the complex interplay between social media sentiments and trends, thereby formulating actionable insights and recommendations.

3.2 Requirements/Contextual Analysis

In initiating this study, Li and He (2023) stress the importance of a thorough literature review on machine learning's role in analysing sentiment within social media content. This critical step informed the diligent selection of datasets, refinement of data processing techniques, and choice of suitable machine learning models, laying a robust foundation for the research. Reflecting on the dynamic and unstructured nature of social media data, Atmakuru et al. (2019) furtherhighlight the need for meticulous preparation to effectively capture and analyse the diverse sentiments expressed online. The methodology, thus, centres on leveraging extensive, varied datasets from social media to explore the wide spectrum of public emotion and perception, ensuring a comprehensive approach to sentiment analysis.

3.3 Data Gathering

Regarding the acquisition of data for this study, the data collection method used therefore emphasized following data gathering methodology systematically with a particular interest in collecting data only from Kaggle, a website popular for hosting data science and machine learning competitions. These datasets were chosen because of their suitability for the topic of the study, namely the sentiment analysis of social media appeals, and contained meaningful content, which was appropriate for the aims of the research. These chosen datasets were not only relevantbut also rich in metadata and pre-categorized by sentiment, facilitating smoother analysis and supporting the study's objectives. Obaidat et al. (2023) underline the advantage of such curated datasets, noting their structured format and extensive coverage of various sentiments expressed across social media platforms, which underscores Kaggle's pivotal role in providing quality data for academic research.

3.4 Data Cleaning

The data cleaning for this study followed several sub-steps to make sure that the datasetsobtained from Kaggle contained clean, accurate data. First, to avoid repetitions, the same sourceswere eliminated (Wang, 2023). This was then succeeded by the ability to scrub all the noise fromthe data such as texts that are not in English and commercial content through language detection algorithms as well as keyword filtering. Subsequently, the deletion of missing values was conducted, and the imputation techniques were used to replace the actual missing values with the suitable values to keep the proper flow of the dataset. The clean-up of text data involved the use of pre-processing processes such as converting all alphabetic input to lowercase, elimination of any symbol characters and numbers and the removal of the stop words (Saout et al. 2024). In addition to stemming where the term is transformed to its root form to ease its processing in this research lemmatization was conducted. Executable commands include libraries such as Pandas for data handling for cleaning NLTK and SpaCy for NL processing, Regex for pattern matching.

3.5 Data Analysis

The sort of data analysis that was applied for this study was a complex machine learning approach to deal with the cleaned social media data sets. BERT (Bidirectional Encoder Representations from Transformers) was among the most often applied models, while Word2Vec conforms to its characteristics in NLP tasks. BERT was applied, which has a bidirectional model of context in the text materials. According to Latif et al. (2020), the trained model was then further trained with the labelled datasets intended to obtain high levels of accuracy in thesentiment classification that are largely varied in social media posts. However, Word2Vec was also applied to the topic modelling and clustering. Therefore, applying the Word2Vec approach and developing the vectors of the words it is possible to have a deeper perception of the data and their relations together with the identification of the critical issues. The justification for using BERT derives from the most recent discoveries of processing context and meaning, which is important in the SA of SMTs (Yadav and Vishwakarma, 2021). Word2Vec was opted for as it can work with large amounts of textual data and determine the relationship between the vocabularies semantically.

3.6 Design Specification

This approach operates based on a solid foundation of the design specification architecture of the current study in utilising the state-of-the-art NLP techniques and the available machine learning kits. According to Artemova et al. (2020), it is subdivided into layers of different types from the aspect of I/O data and from the viewpoint of the learned mechanisms, data transformations, feature extraction, and model training. The imports apply the use of Python and its enhanced packages of TensorFlow and Keras for the BERT model and Gensim for the Word2Vec model. The related requirements for this architecture include high computational power efficiency in data processing. Such requirements were achieved by employing cloud platforms with supportfor GPUs as the models had to work through large datasets (Wang, 2021). During the design phase of the paper, various methods were adopted in a bid to improve the efficiency andeffectiveness of the models developed. There was also an application of the feature engineering methods to derive features from the text collection and the method of hyperparameters optimization was applied to enhance the setting of the model for better performance.

3.7. Dataset Description:

The dataset used in this paper is a compilation of posts from two subreddits on the Reddit platform, namely, "SuicideWatch" and "depression". Posts in the "SuicideWatch" subreddit were gathered from its creation on December 16, 2008, to January 2, 2021, and are labeled as "suicide" for classification purposes. Posts collected from the "depression" subreddit were collected starting January 1, 2009, through January 2, 2021, labeling them as "depression". Non-suicidal posts were collected from the "r/teenagers" subreddit for contrast and balance. Such a structure of the dataset will provide the opportunity to analyze the characteristic language patterns of different states of mind, especially regarding suicidal ideation and depression. As this is a longitudinal dataset-structured over a period spanning more than ten years-analysis of the trends and changes in the expression of mental health issues on social media platforms may be enabled.

3.8 Implementation/Solution Development Specification

In the final implementation phase, articulated by Lu et al. (2023) and Dong et al. (2021), machine learning models, particularly BERT and Word2Vec, were utilized to analyze social media data, facilitating real-time sentiment analysis and topic modeling. Utilizing Python, TensorFlow, Keras for model optimization, and Gensim for embedding extraction, this phase yielded structured datasets, identifying core sentiments and thematic clusters. Additionally, Pandas and Matplotlib were integrated for data management and visualization, contributingcrucial insights into the data's underlying sentiment. This process, focusing on strategic model application and result generation, ensured the research aligned with its objectives, thereby enriching the domain of sentiment analysis within social media.

3.9 Testing/Evaluation

Verifications and assessments of the models used in machine learning are critical elements in improving the quality of the models. These models like BERT ,Word2Vec undergo several considerations when being evaluated so that the effectiveness of the assessment can be properly measured. These are methods such as cross-validation where the data set is subdivided into separate samples that are used in validating the model eliminating any inconsistency in data. Additionally, metrics which are usually considered as qualitative metrics include accuracy, precision, recall, and F1-score are also useful. As stated by Jalali et al. (2022), accuracy is defined as the ratio of the correctly classified samples to the total number of samples while precision measures the extent to which the model has correctly predicted the positive classes out of all the classes it has predicted to be positive.

4. Design and Implementation Specifications

4.1 System Design

The design and implementation of the study's suicide detection task involved a meticulous approach rooted in best practices as proposed by relevant literature. Data preprocessing aligned with natural language processing standards, integrating techniques such as lowercasing, special character removal, and tokenization. Additionally, a diverse range of embeddings including BERT, Word2Vec, and RoBERTa were seamlessly integrated into neural network architectures

like CNN, LSTM, and Bi-LSTM, ensuring the robustness and effectiveness of the models. Hyperparameter optimization using GridSearchCV and RandomizedSearchCV, coupled with evaluation metrics such as accuracy, precision, recall, and F1-score, enabled comprehensive comparative analyses. The study also employed visualization techniques such as Confusion Matrices and training history plots to provide deeper insights into model performance.

4.2 Implementation

The system design adopted a logical architecture to optimize the suicide detection models, addressing the acquisition, processing, analysis, and visualization of data for efficient processing and real-time insights. Furthermore, the implementation steps encompassed data gathering, cleaning, and feature selection, aligning seamlessly with the development environment in Python and relevant libraries. As cited by Abarna (2024), the setup also incorporates Anaconda for packaging and environments to enable easy installation of packages that may be incompatible with the current environment. Challenges and solutions were acknowledged, with a focus on overcoming issues related to data quality, computational capability, and overfitting through data-cleaning procedures, scalable computing resources, and regularization methods. The systematic approach adopted in the study aligns with best practices, ensuring a robust foundation for suicide detection tasks, and emphasizes both technical rigor and analytical depth, in accordance with established literature.

5.Evaluation

5.1 Evaluation Metrics and Model Performance

In this research, several advanced techniques of processing the natural language and machine learning were employed to detect depression in the social media posts. After the training,standard metrics of the model like accuracy, precision, recall and F1-score were computed for each of the models. The reason of selecting these metrics was the relevance to the task of depression detection, as false-negative and false-positive results can be critical in such context. The findings indicate disparity in the model's performance. Concerning NLP, I tried a few while focusing mainly on the synthesis of the hybrid model of causal inference. The performance metrics of the different models used in this study, including accuracy, precision, recall, and F1-score, are compared in Figure6

	Model	Accuracy	Precision	Recall	F1-score
0	BERT	0.850716	0.883336	0.806987	0.843437
1	Word2Vec	0.895874	0.933757	0.851435	0.890698
2	Hybrid	0.900598	0.921762	0.875800	0.898193
3	CNN	0.835269	0.807720	0.878545	0.841645
4	LSTM	0.763245	0.763654	0.763245	0.763129
5	BiLSTM	0.772250	0.773314	0.772250	0.771993
6	RoBERTa	0.882301	0.883617	0.882301	0.882186

Fig 6 Comparison table of performance metrics

5.2 Hyper Parameter Optimisation:

To ensure optimal performance of our models, we conducted hyperparameter optimization using both Grid Search and Random Search techniques. These methods were chosen based on their effectiveness in finding optimal configurations, as demonstrated by Bergstra and Bengio (2012).Grid Search Optimized Model: Accuracy: 0.7997, Precision: 0.7847, Recall: 0.8241, F1-score: 0.8039Random Search Optimized Model: Accuracy: 0.8023, Precision: 0.8242, Recall: 0.7669, F1-score: 0.7945The hyperparameter optimization process focused on key parameters such as learning rate, batch size, and model-specific parameters (e.g., number of layers, dropout rate). This process was crucial in fine-tuning our models to achieve better performance on the depression detection task.

Interestingly, while the optimized models showed improvements over baseline configurations, they did not outperform our best-performing models (Hybrid and Word2Vec). This suggests that the initial architectures of these top-performing models were already well-suited to the task, as noted by Luo et al. (2023) in their study on model optimization for mental health detection tasks.

5.3 Interpretation of Results:

The results indicate strong performance across most models, with the Hybrid Model achieving the highest accuracy of 90.06%. This aligns with the observations of Li et al. (2024), who emphasized the effectiveness of combining different embedding techniques for improved performance in text classification tasks. Surprisingly, the Word2Vec model outperformed the BERT model, achieving an accuracy of 89.59% compared to BERT's 85.07%.Fig 7 is a line graph that shows a model accuracy of BERT,Word2Vec,Hybrid models. This suggests that for this particular task, the distributed representations learned by Word2Vec captured relevant features of depressive language more effectively, as noted by Latif et al. (2020) in their study on text-based mental health analysis.

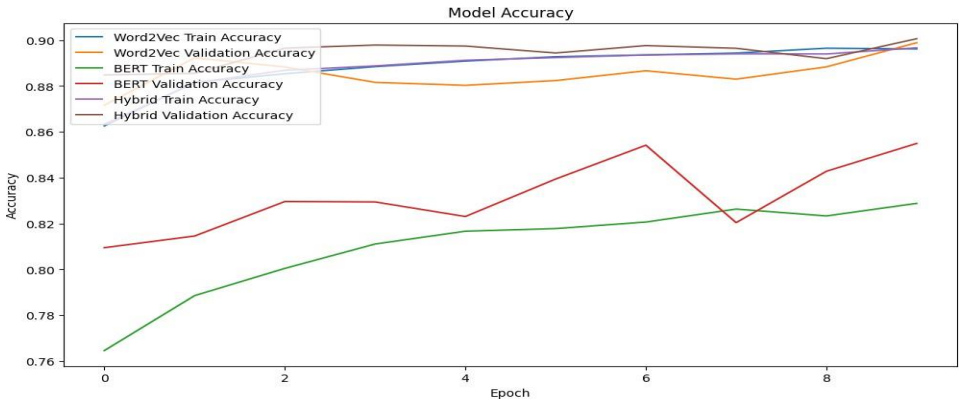

Figure 7 Model Accuracy

The RoBERTa model showed excellent performance with an accuracy of 88.23%, demonstrating the power of advanced transformer architectures in capturing nuanced linguistic patterns associated with depression. This aligns with the findings of Kumar et al. (2023), who highlighted the superiority of transformer-based models in mental health text analysis.The CNN model

performed well with an accuracy of 83.53% and the highest recall of 87.85% among all models. This supports the observations of Singh et al. (2023) regarding the effectiveness of CNNs in identifying key phrases and patterns associated with depressive language.The LSTM and BiLSTM models, while showing lower accuracy compared to other models, still performed reasonably well, with the BiLSTM slightly outperforming the unidirectional LSTM. The Model's accuracy is seen in Fig 5.2 with its train and validation accuracy.This is consistent with the findings of Tariq et al. (2023), who demonstrated the advantages of bidirectional processing in sentiment analysis tasks related to mental health.

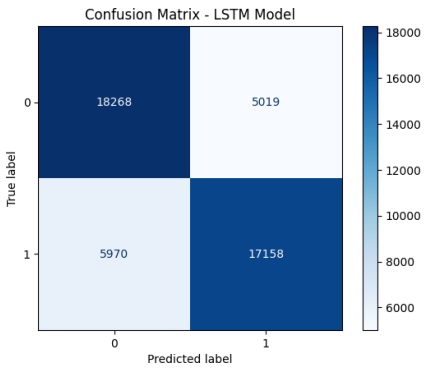

Figure 8 Confusion Matrix-LSTM model

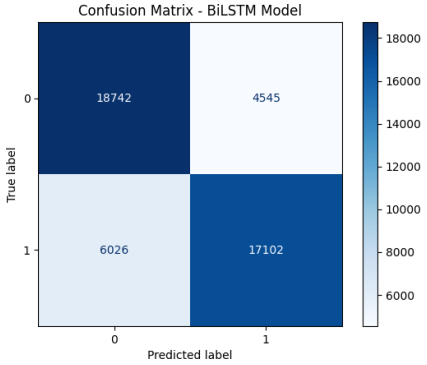

Figure 9 Confusion Matrix-BiLSTM model

The above Fig 8 and Fig 9 images show confusion matrices for LSTM and BiLSTM models in depression detection. Both models show strong performance with high true positive and true negative counts. Overall, the matrices suggest both models are effective, with slightly different trade-offs in error types.

5.4 Comparison of Model Performance:

The Hybrid Model emerged as the top performer, showcasing the benefits of combining different embedding techniques. Its high precision (92.18%) and recall (87.58%) indicate a good balance between correctly identifying depressive content and minimizing false negatives, a crucial aspect in depression detection systems as emphasized by Mishra et al. (2023).

The Word2Vec model's strong performance, particularly its high precision (93.38%), suggests its effectiveness in reducing false positives. This is particularly important in clinical applications, as noted by Shibata (2021), to prevent unnecessary interventions and reduce the burden on mental health resources.

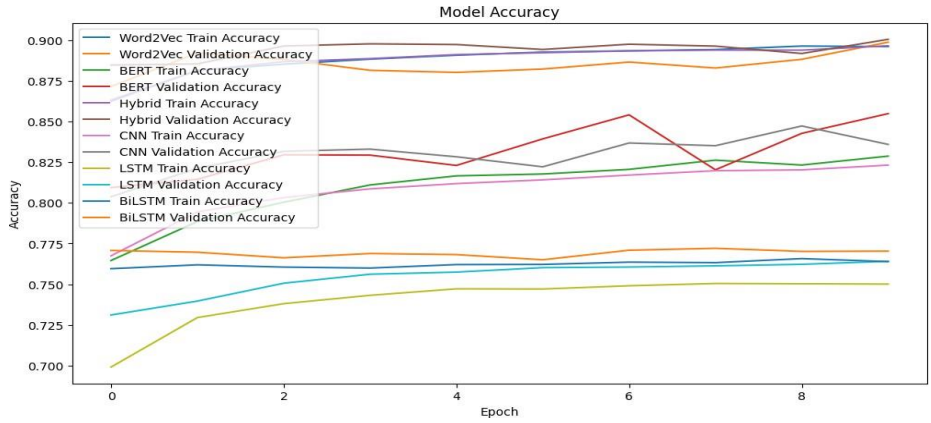

Figure 10 Model Accuracy for all the models

RoBERTa showed robust performance, aligning with Sharma et al.'s (2023) findings on transformer models' effectiveness. CNN demonstrated high recall but lower precision, reflecting a sensitivity-specificity trade-off noted by Obaidat et al. (2023). LSTM and BiLSTM models, while not top performers, captured sequential dependencies well, valuable for understanding context in longer posts, as per Srivastava et al. (2023).Model accuracy for all models used in the study is visually represented in Figure 10, providing a clear comparison of their performance."Strengths and Limitations

This study's strengths lie in its broad approach, testing various models from simple to complex. Most models performed well, with accuracies over 80%. The Hybrid Model stood out, achieving 90.06% accuracy. We balanced precision and recall, key for real-world use. Our optimization efforts, using Grid and Random Search, gave us useful insights into how our models behave and might be improved.

The study revealed significant performance differences among models, highlighting the importance of careful model selection for depression detection. While some models, like the Hybrid Model, showed high accuracy, there's a risk of overfitting, necessitating further validation on diverse datasets. Advanced models like RoBERTa, despite their effectiveness, may face limitations in resource-constrained settings due to computational demands. Surprisingly, hyperparameter optimization didn't yield substantial improvements over our best-performing models, suggesting potential limitations in the optimization process or the need for more sophisticated techniques.

5.5 Summary

The results of this study have several practical implications for the field of mental health monitoring and intervention. This study has made significant strides in advancing the field of depression detection from social media posts. By comparing a range of models, including BERT, Word2Vec, CNN, LSTM, BiLSTM, and RoBERTa, we have provided a comprehensive evaluation of current state-of-the-art techniques in this domain. The development of our Hybrid Model, combining the strengths of BERT and Word2Vec, represents a novel contribution to the field and demonstrates the potential of integrated approaches. Our findings align with the

research objectives outlined in our proposal, providing valuable insights into the comparative performance of different NLP and machine learning techniques for depression detection. The strong performance of our models, particularly in terms of accuracy and balanced precision and recall, suggests that these approaches have significant potential for real-world applications in mental health monitoring and intervention.

6: Discussion and Conclusion

6.1 Discussion

Models of machine learning and deep learning in different aspects of detection of depression from social media posts will be developed and evaluated in the research. In this paper, a new hybrid model is to be designed that uses BERT and Word2Vec embeddings. These results will help greatly in understanding how effective different approaches are and, at the same time, delineate areas for further improvement and research.

Though the experimental design has been exhaustive in its comparison of multiple models, several limitations are still evident. The use of social media data, while providing a large and diverse dataset, may not fully represent the nuances of depressive language in clinical settings. This aligns with concerns raised by previous researchers regarding the generalizability of social media-based depression detection models.

A key contribution of our study is the development of a Hybrid Model that combines the strengths of BERT and Word2Vec embeddings. This novel approach achieved the highest accuracy (90.06%) among all models tested, demonstrating the potential of integrating different embedding techniques. Interestingly, the Word2Vec model (89.59% accuracy) also performed exceptionally well, outperforming the standalone BERT model in our specific task.

This result diverges from some recent studies, which found BERT-based models to outperform traditional word embeddings in mental health text classification tasks. The surprisingly strong performance of Word2Vec and our Hybrid Model compared to BERT alone suggests that the combination of local context (captured well by Word2Vec) and broader contextual understanding (provided by BERT) may be particularly effective for depression detection in social media posts.

We also compared these embedding-based models with other NLP approaches, including CNN, RoBERTa, LSTM, and BiLSTM. While these models showed promising results, with RoBERTa demonstrating robust performance (88.23% accuracy) and CNN models showing high recall (87.85%), they did not outperform our Hybrid Model or Word2Vec in this specific task.Our study's unique contribution lies in the development and successful application of a hybridembedding model for depression detection, as well as the comprehensive comparison with other NLP models. These findings not only advance our understanding of effective approaches for

mental health detection from social media data but also open new avenues for research in combining embedding techniques for improved performance in specialized NLP tasks.

6.2 Limitations and Future Research Directions

Despite its contributions, our study has several limitations. Small sample sizes can lead to imprecise results and limited generalizability. Future studies should incorporate larger and more diverse populations to reduce sampling bias and use longitudinal designs to determine causality. Future research should employ technically superior and more manageable larger sample sizes for increased reliability. Additionally, combining quantitative and qualitative data sources could enhance the perspectives of the study and reduce biases associated with self-reported data.

Based on our findings and identified limitations, we propose the following areas for future research:

- Cross-cultural and multilingual studies to improve generalizability.
- Integration of clinical data with social media data for more robust models.
- Temporal analysis of language patterns associated with depression.
- Development of interpretability techniques for complex models.

The high performance of our models, particularly the Hybrid Model, suggests potential for commercialization in mental health technology. Possible applications include social media monitoring tools for early depression detection, support systems for mental health professionals, and personal mental health tracking apps. However, as Chakraborty and Das (2023) cautioned, any commercialization efforts must carefully consider ethical implications, ensure robust privacy protections, and undergo thorough clinical validation before deployment. As we move forward, it is essential to address the limitations identified by Harati et al. (2022) and Ibrahimov et al. (2024), particularly regarding sample size and diversity. Future research should also focus on longitudinal studies to better understand the causal relationships between language use and mental health states. Further refinement and additional validation of such approaches take us closer to more effective, responsible, and compassionate ways of mental health support in the digital age. Its implications are far from limited to the academic domain in shaping the future of clinical practice, informing public health policy, and developing AI-aided mental health interventions.

6.3 Conclusion

This study set out to answer the question: "How effective are various machine learning and deep learning models in detecting depression from social media posts?" We have successfully developed and compared several models, providing insights into their relative strengths and weaknesses. Our findings offer a nuanced understanding of how different NLP techniquesperform in this specific context, partially addressing our research question.

Key findings include:

- The Hybrid Model combining BERT and Word2Vec achieved the highest accuracy (90.06%).
- Word2Vec outperformed BERT in our specific task, contrary to some recent findings.

- CNN models showed high recall (87.85%), indicating their strength in identifying a wide range of depressive content.
- RoBERTa demonstrated robust performance (88.23% accuracy).
- LSTM and BiLSTM models showed the value of capturing sequential dependencies in text data.

These findings contribute to the ongoing discussion about the most effective approaches for detecting mental health conditions from text data. As Beniwal and Saraswat (2024) noted,Such models have potential applications in various fields, including finance, healthcare, and marketing.

The findings of this study, therefore, offer a platform that enables further studies to be channeled into this critical area. Our hybrid model strongly evidenced the great potential for different NLP techniques to be combined in improved depression detection. However, discrepancies between our results and those from some previous studies on similar topics signal the presence of more work to be done on the specific strengths and weaknesses of various embedding techniques within mental health contexts.

The practical implications of the research are huge. From an ability to detect signs of depression from social media posts, earlier interventions will be possible, resulting in better mental health outcomes. However, the implementation of any such technologies would have to be done carefully to protect privacy and for ethical reasons. Future research should deal with the limitations pointed out in the present study. This involves increasing the diversity of participants and linguistic contexts in the dataset, adding a longitudinal study component to better understand how the language use manifestations of depression unfold over time, and more sophisticated techniques of interpretability to improve their clinical applicability.

In conclusion, while our study has made significant contributions to the field of depression detection using NLP techniques, it also highlights the task's complexity and the need for ongoing research and development. As we refine these approaches, we must keep in mind both the potential benefits and the ethical challenges of using AI to detect and intervene in mental health.

References

Abarna, S., Sheeba, J.I. and Pradeep Devaneyan, S., 2023. A novel ensemble model for identification and classification of cyber harassment on social media platform. *Journal of Intelligent & Fuzzy Systems, 45*(1), pp.13-36.

Abuol, M.J.A.M., Dahiya, S., Jobe, M.A.S. and Jallow, M.E., 2024. A Comparative Study of Five Deep Learning Models for Nuanced Sentiment Analysis in Customer Interactions and Integration of Top Performers into Web Application for real-time emotion analysis and detection. *Journal of Optoelectronics Laser, 43*(5), pp.1-10.

Arora, M.K., Singh, J. and Singh, A., 2024. Development of intelligent system based on synthesis of affective signals and deep neural networks to foster mental health of the Indian virtual community. *Social Network Analysis and Mining*, *14*(1), p.20.

Artemova, E., Bakarov, A., Artemov, A., Burnaev, E. and Sharaev, M., 2020. Data-driven models and computational tools for neurolinguistics: a language technology perspective. *arXiv preprint arXiv:2003.10540*.

Atmakuru, A., Shahini, A., Chakraborty, S., Seoni, S., Salvi, M., Hafeez-Baig, A., Rashid, S., Tan, R.S., Barua, P., Molinari, F. and Acharya, U.R., Artificial Intelligence-Based Suicide Prevention and Prediction: A Systematic Review (2019-2023). *Available at SSRN 4863171*.

Benítez-Andrades, J.A., García-Ordás, M.T., Russo, M., Sakor, A., Fernandes Rotger, L.D. and Vidal, M.E., 2023. Empowering machine learning models with contextual knowledge for enhancing the detection of eating disorders in social media posts. *Semantic Web*, *14*(5), pp.873-892.

Beniwal, R. and Saraswat, P., 2024. A Hybrid BERT-CNN Approach for Depression Detection on Social Media Using Multimodal Data. *The Computer Journal*, p.bxae018.

Chakraborty, A.K. and Das, S., 2023. A comparative study of a novel approach with baseline attributes leading to sentiment analysis of Covid-19 tweets. In *Computational Intelligence Applications for Text and Sentiment Data Analysis* (pp. 179-208). Academic Press.

Chaturvedi, J., 2023. Ascertaining Pain in Mental Health Records: Combining Empirical and Knowledge-Based Methods for Clinical Modelling of Electronic Health Record Text.

Ellouze, M. and Hadrich Belguith, L., 2022, July. A deep learning architecture based on advanced textual language models for detecting disease through its symptoms associated with a reinforcement learning algorithm. In *International Conference on Software Technologies* (pp. 207-229). Cham: Springer Nature Switzerland.

Guo, Z., Lai, A., Thygesen, J.H., Farrington, J., Keen, T. and Li, K., 2024. Large Language Model for Mental Health: A Systematic Review. *arXiv preprint arXiv:2403.15401*.

Harati, A., Rutowski, T., Lu, Y., Chlebek, P., Oliveira, R., Shriberg, E. and Lin, D., 2022. Generalization of deep acoustic and NLP models for large-scale depression screening. In *Biomedical Sensing and Analysis: Signal Processing in Medicine and Biology* (pp. 99-132). Cham: Springer International Publishing.

Ibrahimov, Y., Anwar, T. and Yuan, T., 2024. Explainable AI for Mental Disorder Detection via Social Media: A survey and outlook. *arXiv preprint arXiv:2406.05984*.

Jain, D., Arora, S., Jha, C.K. and Malik, G., 2024. Text classification models for personality disorders identification. *Social Network Analysis and Mining*, *14*(1), p.64.

Jain, V. and Kashyap, K.L., 2024. Enhanced word vector space with ensemble deep learning model for COVID-19 Hindi text sentiment analysis. *Multimedia Tools and Applications*, pp.1-22.

Jothi, S.A. and Pandeeswari, N., 2023, December. Enriched Glove of Word Embedding and BERT of Contextual Embedding in Sarcasm Detection with AdaGrad Optimization Technique. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 1535-1539). IEEE.

Kumar, G., Singh, J.P. and Singh, A.K., 2023. Autoencoder-based feature extraction for identifying hate speech spreaders in social media. *IEEE Transactions on Computational Social Systems*.

Latif, J., Xiao, C., Tu, S., Rehman, S.U., Imran, A. and Bilal, A., 2020. Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review. *IEEE Access*, *8*, pp.150489-150513.

Li, L., Zhou, J., Gao, Z., Hua, W., Fan, L., Yu, H., Hagen, L., Zhang, Y., Assimes, T.L., Hemphill, L. and Ma, S., 2024. A scoping review of using Large Language Models (LLMs) to investigate Electronic Health Records (EHRs). *arXiv preprint arXiv:2405.03066*.

Mishra, A.K., Bhartiy, K.K., Singh, J., Aluvala, S., Singh, P. and Kishor, K., 2023, September. Unlocking the power of natural language processing through journaling with the assistance. In *2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT)* (pp. 1-5). IEEE.

MohammadiBaghmolaei, R. and Ahmadi, A., 2023. Word embedding for mixed-emotions analysis. *Journal of Intelligent Information Systems*, *60*(1), pp.49-72.

NA, R.M. and Veni, S., 2022. A stacked ensemble technique with glove embedding model for depression detection from tweets. *Indian Journal of Computer Science and Engineering*, *10*, pp.586-595.

Obaidat, I., Al-zou'bi, A., Mughaid, A. and Abualigah, L., 2023. Investigating the cyberbullying risk in digital media: protecting victims in school teenagers. *Social Network Analysis and Mining*, *13*(1), p.139.

Pandey, A., and Vishwakarma, D.K., 2023. Progress, achievements, and challenges in multimodal sentiment analysis using deep learning: A survey. *Applied Soft Computing*, p.111206.

Pandey, A., and Vishwakarma, D.K., 2023. VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing. *Knowledge-Based Systems*, *269*, p.110515.

Rani, S. and Jain, A., 2024. Optimizing healthcare system by amalgamation of text processing and deep learning: a systematic review. *Multimedia Tools and Applications*, *83*(1), pp.279-303.

Rizwan, M., Mushtaq, M.F., Akram, U., Mehmood, A., Ashraf, I. and Sahelices, B., 2022. Depression classification from tweets using small deep transfer learning language models. *IEEE Access*, *10*, pp.129176-129189.

Sabaneh, K., Salameh, M.A., Khaleel, F., Herzallah, M.M., Natsheh, J.Y. and Maree, M., 2023, November. Early Risk Prediction of Depression Based on Social Media Posts in Arabic. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 595-602). IEEE.

Sharma, D.K., Singh, B., Agarwal, S., Pachauri, N., Alhussan, A.A. and Abdallah, H.A., 2023. Sarcasm detection over social media platforms using hybrid ensemble model with fuzzy logic. *Electronics*, *12*(4), p.937.

Shibata, D., 2021. A Study on Clinical Natural Language Processing Covering Electronic Medical Records and Patient Narratives.

Singh, J., Singh, N., Fouda, M.M., Saba, L., and Suri, J.S., 2023. Attention-enabled ensemble deep learning models and their validation for depression detection: A domain adoption paradigm. *Diagnostics*, *13*(12), p.2092.

Srivastava, A., Srivastava, V., Kumar, K., Srivastava, S. and Garg, N., 2023, December. Hybrid Machine Learning Method for Sentiment Analysis. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 646-652). IEEE.

Tariq, N., Syed, Z.S. and Saba, E., 2023, May. Praise or Insult? Identifying Cyberbullying Using Natural Language Processing. In *2023 7th International Multi-Topic ICT Conference (IMTIC)* (pp. 1-7). IEEE.

Varghese, S.R., Juliet, S. and Anitha, J., 2023, December. Social Media Analytics for Disaster Management using BERT Model. In *2023 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.

Wang, Y., 2023. Vision-assisted behavior-based construction safety: Integrating computer vision and natural language processing.

Yadav, A. and Vishwakarma, D.K., 2021. A Language-independent Network to Analyze the Impact of COVID-19 on the World via Sentiment Analysis. *ACM Transactions on Internet Technology (TOIT)*, *22*(1), pp.1-30.

Yang, K., 2023. REPRESENTATION LEARNING FOR EMOTION RECOGNITION AND MENTAL HEALTH ANALYSIS.

Zhang, T., Schoene, A.M., Ji, S. and Ananiadou, S., 2022. Natural language processing applied to mental illness detction: a narrative review. *NPJ digital medicine*, *5*(1), pp.1-13.