Contrastive Learning and Large Language Models for Depression Detection from Social Media

By

Anay Sinhal

Under The Supervision of

Dr. Bhavesh N. Gohil

Associate Professor, Deptt. of Computer Sci. & Engg.,

Sardar Vallabhbhai National Institute of Technology,

Surat (Gujarat)

July 2024

DECLARATION

I hereby declare that the research work titled "**Contrastive Learning and Large Language Models for Depression Detection from Social Media**" is the result of my own research conducted under the supervision of **Dr. Bhavesh N. Gohil** at Sardar Vallabhbhai National Institute of Technology, Surat.

This work, completed with over 8 weeks of dedicated research and engagement, has not been submitted for any degree or examination in this or any other institution.

I affirm that all sources of material used for the thesis have been duly cited, ensuring the originality and integrity of the research.

Anay Sinhal

ACKNOWLEDGEMENT

I am profoundly grateful for the guidance, support, and expertise bestowed upon me by Dr. Bhavesh N. Gohil, my esteemed project supervisor, whose insights and encouragement were pivotal to the trajectory and success of my research on Depression Detection using Contrastive Learning. His mentorship was invaluable, and I extend my heartfelt thanks.

I am indebted to my college's administrative and academic leadership, including Prof. Dheeraj Sanghi, the Vice Chancellor; Prof. Renu Jain, the Dean; and our Department Coordinators, Prof. Amit Sinhal and Prof. Taruna Sunil. Their unwavering support and the provision of an enriching academic environment have been crucial to my research endeavours.

Special thanks are extended to my peers and colleagues, whose collaborative spirit, stimulating discussions, and shared passion for research enriched this project in countless ways. Their camaraderie and intellectual generosity have been a source of inspiration and motivation throughout this journey.

The resources and facilities provided by JK Lakshmipat University were indispensable to the execution and completion of this study. For this support, I am incredibly grateful.

Finally, I must acknowledge my family and friends' enduring patience, love, and encouragement. Their belief in my capabilities and unwavering support gave me the strength and perseverance required to pursue and achieve my academic goals.

This project stands as a testament to the collective effort, encouragement, and guidance of everyone mentioned above, and to all of them, I offer my deepest gratitude. Their contributions have facilitated the successful completion of this research and profoundly impacted my personal and professional growth.

Anay Sinhal

TABLE OF CONTENTS

ABSTRACT

Depression is a pervasive mental health disorder that significantly impacts individuals worldwide, often manifesting in subtle and complex ways. With the proliferation of social media, platforms like Reddit and Twitter have become critical venues for expressing depressive symptoms, providing a rich source of data for mental health analysis. This research explores the use of contrastive learning with large language models (LLMs) for the detection of depression, leveraging textual data from social media to enhance early identification and intervention strategies.

We used the DepressionEmo dataset, comprising annotated Reddit posts reflecting various depression-related emotions. This study employs advanced computational techniques, including text-based analysis, natural language processing (NLP), and machine learning (ML) models, to detect depressive symptoms. Key models such as Support Vector Machines (SVM), LightGBM, and XGBoost are implemented alongside deep learning models like GAN-BERT, BERT, and BART. The research emphasizes contrastive learning to improve the detection accuracy of these models by distinguishing between similar and dissimilar data points.

Our methodology involves comprehensive data augmentation techniques to enhance the robustness and diversity of the dataset. The models are trained and fine-tuned on the DepressionEmo dataset, with performance evaluated using precision, recall, and F1-score metrics. The results demonstrate significant improvements in the accuracy and scalability of depression detection models, highlighting the potential of automated systems for real-time mental health monitoring and intervention.

This study addresses critical research gaps, such as data diversity, subtle contextual indicators, and the need for explainable AI in clinical applications. By integrating multimodal data and ensuring ethical considerations in data usage, the proposed models contribute to the development of reliable and scalable solutions for global mental health challenges.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Overview

Depression is a widespread mental health disorder affecting millions globally, with social media platforms like Reddit and Twitter offering valuable data for detecting depressive emotions. This report focuses on using contrastive learning with large language models (LLMs) to enhance early and accurate detection of depression.

1.2 Depression

Depression, marked by persistent sadness and loss of interest, is a leading cause of disability worldwide. Traditional diagnostic methods are often subjective and time-consuming, driving interest in automated, data-driven approaches. Machine learning and natural language processing (NLP), particularly through social media analysis, offer promising solutions. By leveraging advanced techniques like contrastive learning and LLMs, researchers can detect subtle linguistic signs of depression, providing scalable and accurate early intervention tools.

1.3 Depression on Social Media

Social media platforms like Reddit and Twitter are key sources for analyzing real-time data to detect depression.

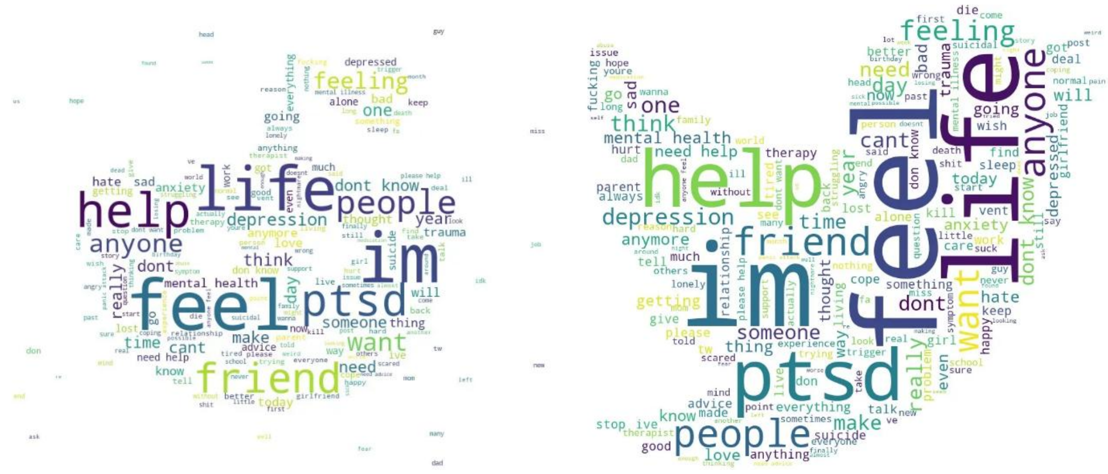

Figure 1.1: Word cloud for depressing texts from Reddit and Twitter

Reddit: Communities like r/depression and r/anxiety provide rich qualitative data where users openly discuss their emotional states, making it a valuable resource for identifying depressive symptoms.

Twitter: With its real-time, brief posts, Twitter offers a continuous stream of textual data reflecting users' mental states. Despite the brevity of tweets, they present opportunities to develop models for quick detection of distress.

Data analysis from these platforms involves steps like data collection, preprocessing, and annotation. The DepressionEmo dataset, derived from Reddit posts, includes depression-related emotion labels annotated using a zero-shot classification method and validated against human and ChatGPT annotations. This robust annotation process is crucial for training reliable depression detection models.

1.4 Depression Detection

Detecting depression involves identifying patterns and indicators in data that correlate with depressive symptoms. Traditional clinical methods rely on structured interviews and self-assessment questionnaires, but these methods are not scalable for large populations. Consequently, there is a shift towards automated systems that can analyze text data from social media to identify signs of depression.

Text-Based Analysis: The primary method for detecting depression in social media posts is text-based analysis. This involves using NLP techniques to extract syntactic and semantic features from text that indicate depressive emotions. Common techniques include sentiment analysis, topic modeling, and keyword extraction.

Machine Learning and Deep Learning Models: These models are employed to classify and predict depression from text data. Models like BERT, GPT-4, and RoBERTa are used for their powerful language understanding capabilities. In particular, the BART model has shown high performance in detecting depression-related emotions, achieving the highest F1-Macro score in the DepressionEmo dataset.

Contrastive Learning: This report focuses on the use of contrastive learning, a technique from self-supervised learning that helps models distinguish between similar and dissimilar data points. By creating augmented versions of text data and forming positive and negative pairs, the model learns to capture subtle nuances in text indicative of depressive emotions. This approach enhances the model's ability to generalize and accurately detect depression.

By integrating these advanced techniques, researchers aim to develop robust and scalable models for detecting depression from social media data, ultimately contributing to better mental health monitoring and intervention strategies.

1.5 Depression Detection Techniques

The detection of depression through automated systems has evolved significantly, incorporating various advanced techniques to improve accuracy and efficiency. The following are key techniques used in depression detection:

Natural Language Processing (NLP): NLP is a fundamental technique for text-based depression detection. It involves the analysis of textual data to extract meaningful information. Techniques such as tokenization, part-of-speech tagging, sentiment analysis, and topic modeling are commonly used to process and analyze text data from social media.

Machine Learning Algorithms: Traditional machine learning algorithms like Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forests have been used to classify and predict depressive symptoms. These models rely on hand-crafted features extracted from text data to make predictions.

Deep Learning Models: Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown superior performance in capturing complex patterns in text data. Models like Long Short-Term Memory (LSTM) networks and BERT (Bidirectional Encoder Representations from Transformers) leverage large-scale pre-training to understand the context and semantics of text, improving the accuracy of depression detection.

Hybrid Models: Hybrid approaches combine multiple techniques to enhance detection accuracy. These models may integrate text, audio, and visual data with advanced machine learning algorithms to provide a comprehensive analysis of depressive symptoms.

1.6 Research Gap

Despite significant advancements in the field of depression detection, several gaps remain that need to be addressed to improve the effectiveness and reliability of these systems:

Limited Data Diversity: Many existing studies rely on datasets from specific social media platforms or regions, which may not capture the full diversity of language and cultural expressions of depression. There is a need for more diverse datasets that represent a broader range of demographics and linguistic variations.

Subtle and Contextual Indicators: Detecting depression involves understanding subtle and context-dependent indicators that may not be easily captured by traditional models. This requires advanced techniques that can accurately interpret the nuances of language and context.

Generalization to Real-World Scenarios: Models trained on specific datasets may not generalize well to different populations or real-world settings. Ensuring that models can perform consistently across various contexts is a critical challenge.

Explainability and Transparency: While deep learning models, including those using contrastive learning, have shown high accuracy, they often lack transparency in their decision-making processes. Developing explainable models that provide insights into their predictions is essential for clinical applications and user trust.

Integration of Multimodal Data: Most current approaches focus primarily on text data, while depression can manifest through various modalities, including voice, facial expressions, and physiological signals. Integrating multimodal data can provide a more comprehensive understanding of an individual's mental state.

Ethical and Privacy Concerns: The use of social media data for depression detection raises ethical and privacy concerns. Ensuring the anonymity and consent of users, as well as addressing potential biases in the data, are important considerations for developing responsible AI systems.

By addressing these research gaps, future work can enhance the accuracy, generalizability, and ethical implications of depression detection systems, contributing to better mental health outcomes on a global scale.

1.7 Problem Statement and Objectives

The increasing prevalence of depression and the limitations of traditional detection methods necessitate the development of advanced, automated systems that can accurately identify depressive symptoms from large volumes of data. This research aims to address these challenges by leveraging contrastive learning with large language models to enhance the detection of depression-related emotions in social media posts.

Traditional methods for detecting depression are often subjective, time-consuming, and not scalable to large populations. There is a need for automated, scalable, and accurate systems that can detect depression from social media data, capturing subtle linguistic nuances and contextual indicators of depressive emotions.

1. Utilize the DepressionEmo dataset derived from Reddit posts, annotated with multiple depression-related emotions using a zero-shot classification method validated against human and ChatGPT annotations.
2. Apply contrastive learning techniques to enhance the model's ability to distinguish between similar and dissimilar text data, thereby improving the detection of nuanced depressive emotions.
3. Utilize pre-trained large language models (LLMs) such as BERT to generate text embeddings, followed by fine-tuning with supervised learning on the labeled dataset to improve classification accuracy.
4. Assess the effectiveness of the depression detection model using comprehensive evaluation metrics such as accuracy, precision, recall and F1-score. Compare the performance with baseline models to demonstrate improvements.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews various approaches and methodologies used in depression detection, highlighting key features, evaluation criteria, and comparisons with existing models.

2.1 Types of Depression Detection Techniques

Depression detection techniques have evolved significantly with advancements in technology. The following subsections categorize these techniques based on the type of data and methods used.

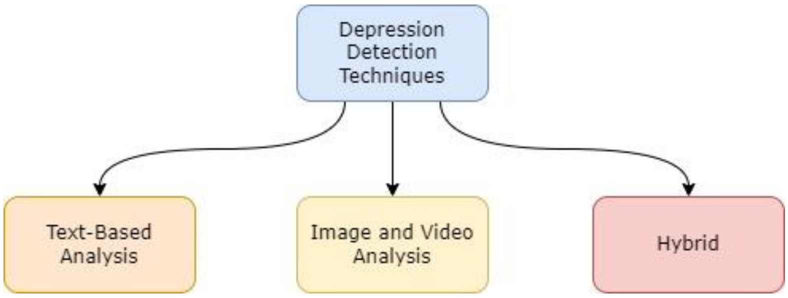

Figure 2.1: Types of Depression Detection Techniques

2.1.1 Text-Based Analysis

Several studies utilize text-based analysis to detect depression by examining the linguistic features of social media posts and other written content. Techniques include:

- Natural Language Processing (NLP): Extracting syntactic and semantic features from text (e.g., "Detecting Depression Using Emotion Artificial Intelligence" by Deshpande & Rao, 2017).
- Sentiment Analysis: Assessing the sentiment of posts to determine depressive tendencies (e.g., "Depression Detection with Sentiment Analysis of Tweets" by Hemanth Kumar M. & Latha A., 2019).

2.1.2 Image and Video Analysis

Some approaches focus on visual data to identify depression-related cues:

- Facial Emotion Recognition: Using image processing techniques to detect facial expressions indicative of depression (e.g., "Image Processing Techniques to Recognize Facial Emotions" by Rani & Durgadevi, 2017).
- Multimodal Analysis: Combining video and audio data for comprehensive emotion detection (e.g., "Detecting Depression Using a Framework Combining Deep Multimodal Neural Networks" by Victor et al., 2019).

2.1.3 Hybrid Approaches

Combining multiple data sources and techniques to improve detection accuracy:

- Hybrid Models: Integrating text, audio, and visual data with advanced ML algorithms (e.g., "Hybrid Machine Learning Models to Detect Signs of Depression" by Khan & Alqahtani, 2024).

2.2 Features and Parameters

Effective depression detection relies on identifying relevant features and parameters. Key features include:

2.2.1 Linguistic Features

- Word Frequency and Sentiment Scores: Analyzing the frequency of depressive words and overall sentiment (e.g., "Predicting Depression Levels Using Social Media Posts" by Aldarwish & Ahmad, 2017).

- Syntax and Grammar: Utilizing advanced NLP techniques to extract meaningful patterns from text.

2.2.2 Visual and Acoustic Features

- Facial Expressions: Detecting changes in facial expressions using image processing algorithms.
- Voice Tone and Pitch: Analyzing speech patterns to identify depressive symptoms.

2.2.3 Behavioral Features

- User Activity Patterns: Monitoring social media activity, such as posting frequency and engagement (e.g., "Machine Learning-based Approach for Depression Detection in Twitter" by AlSagri & Ykhlef, 2020).

2.3 Evaluation Criteria

Evaluation of depression detection models is critical to ensure their effectiveness and reliability. Common evaluation criteria include:

2.3.1 Accuracy

The percentage of correctly identified cases of depression out of the total cases.

2.3.2 Precision and Recall

- Precision: The proportion of true positive results among all positive results.
- Recall: The ability of the model to identify all relevant instances of depression.

2.3.3 F1-Score

The harmonic mean of precision and recall, providing a single metric to evaluate model performance.

2.3.4 Specificity and Sensitivity

- Specificity: The ability to correctly identify non-depressive instances.
- Sensitivity: The ability to correctly identify depressive instances.

2.4 Comparison with Existing Models

Comparing new models with existing ones highlights advancements and areas for improvement. Key points of comparison include:

2.4.1 Traditional Models

Naive Bayes and SVM: Earlier models primarily used for text classification and sentiment analysis.

2.4.2 Deep Learning Models

CNN and LSTM: Advanced models that improve accuracy by capturing complex patterns in data (e.g., "Predicting Depression Using Deep Learning and Ensemble Algorithms on Raw Twitter Data" by Shetty et al., 2020).

2.4.3 Hybrid Approaches

Combining Multiple Techniques: Hybrid models that integrate various data sources and methods to enhance performance (e.g., "Multi Class Depression Detection Through Tweets using Artificial Intelligence" by Nusrat et al., 2024).

2.5 A Review of the Work Done in the Field

The following reviews provide a comprehensive survey about the developments in the depression detection technology around the world.

Multi-Class Depression Detection Through Tweets using AI by **Muhammad Osama Nusrat et al. [1]** (2024) explores detecting five types of depression (bipolar, major, psychotic, atypical, postpartum) from Twitter data using explainable AI. The study fine-tunes a BERT model on annotated tweets, achieving 96% accuracy, surpassing other models. This work underscores the importance of explainable AI in enhancing the reliability of depression detection on social media.

Explainable Depression Detection Using Large Language Models on Social Media Data by **Yuxi Wang et al. [2]** (2024) presents a method to detect depression by analyzing Reddit posts with large language models (LLMs). The approach automates Beck's Depression Inventory (BDI) scoring using LLMs, outperforming state-of-the-art methods in various metrics. The research highlights LLMs' potential in mental health monitoring, emphasizing transparent explanations crucial for clinical use and trust.

Hybrid Machine Learning Models to Detect Signs of Depression by **Shakir Khan et al. [3]** (2024) investigates hybrid models for depression detection via sentiment analysis of Twitter tweets. The study compares four models, with the best achieving 99.4% accuracy. The research suggests integrating hybrid machine learning models into mental health monitoring systems for early depression detection and intervention.

Depression Detection from Social Media Text Analysis using NLP Techniques and Hybrid Deep Learning Model by **Vankayala Tejaswini et al. [4]** (2024) enhances depression detection by combining Fasttext embeddings with CNN and LSTM networks. Applied to Reddit and Twitter datasets, the model shows superior accuracy over traditional methods, suggesting potential for automating medical diagnosis systems and developing mental health chatbots.

Improving Disease Detection from Social Media Text via Self-Augmentation and Contrastive Learning by **Pervaiz Iqbal Khan et al. [5]** (2024) introduces a method combining Contrastive Learning with self-augmentation to enhance disease detection from social media posts. The approach shows up to a 2.48% improvement in F1-score, highlighting its potential for public health monitoring and policy-making.

Deep Learning for Depression Detection Using Twitter Data by **Doaa Sami Khafaga et al. [6]** (2023) enhances depression detection accuracy via a Multi-Aspect Depression Detection with Hierarchical Attention Network (MDHAN) combined with Adaptive Particle Swarm Optimization and Grey Wolf Optimization (MDH-PWO) for feature selection. Using over 4,000 tweets from Kaggle, the MDH-PWO model achieves 99.86% accuracy, outperforming traditional methods like CNN and SVM.

Depression Detection in Social Media Comments Data Using Machine Learning Algorithms by **Zannatun Nayem Vasha et al. [7]** (2023) focuses on detecting depression through machine learning analysis of 10,000 comments from Facebook and YouTube. After pre-processing and TF-IDF vectorization, six classifiers were tested, with the SVM achieving the highest accuracy of 75.15%.

Survey on Design Efficient Depression Detection Using Machine Learning by **Mayuri P. Saraf et al. [8]** (2023) reviews machine learning applications for detecting depression from various data sources, including social media and medical records. The study discusses the use of supervised and unsupervised methods for early detection and intervention, emphasizing the potential of machine learning to enhance diagnosis, personalized treatment, and management of depression, especially in underdiagnosed populations.

Depression Detection Using Emotional Artificial Intelligence and Machine Learning: A Closer Review by **Manju Lata Joshi et al. [9]** (2022) provides a comprehensive review of AI and ML techniques for depression detection, focusing on text and image processing, and chatbots. The paper highlights that Multinomial Naive Bayes often outperforms other models in text-based sentiment analysis of tweets.

AI Therapist Using Natural Language Processing by **Shephali Santosh Nikam et al. [10]** (2020) aims to develop an AI-based system for managing stress through NLP, targeting youth and IT professionals. The system, built with Python, Flask, and Chatterbot, uses collaborative filtering and Naive Bayes for emotion classification. The chatbot demonstrated improved accuracy and response time, offering tailored content based on the user's emotional state.

Depression Detection of Tweets and A Comparative Test by **P. V. Rajaraman et al. [11]** (2020) explores depression detection from tweets using various machine learning and deep learning algorithms. The study employs Sentiment 140, TWINT tweet scraps, and Google Word2Vec data, utilizing models like Naive Bayes, Linear Support Vector, Logistic Regression, TF-IDF, and LSTM. The LSTM model achieved the highest accuracy in detecting depressive tweets, with the study suggesting future work on other mental health issues like PTSD, stress, and anxiety.

Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features by **Hatoon S. AlSagri et al. [12]** (2020) investigates depression detection by analyzing tweets and user behavior on Twitter. Using data from 111 user profiles and over 300,000 tweets, the study employs classifiers like SVM, Naive Bayes, and Decision Tree, finding that SVM with a linear kernel provides the highest accuracy (82.5%).

Predicting Depression Using Deep Learning and Ensemble Algorithms on Raw Twitter Data by **Nisha P. Shetty et al. [13]** (2020) examines the use of deep learning and ensemble algorithms to detect depression in Twitter users. The study utilizes an LSTM model and a CNN model, with the CNN achieving a higher test accuracy (95%). Logistic Regression with TF-IDF vectorizer was the best-performing traditional classifier.

Depression Detection with Sentiment Analysis of Tweets by **Hemanthkumar M. et al. [14]** (2019) applies NLP techniques for sentiment analysis of tweets to detect depression. Using 43,000 tweets from Kaggle, the study employs Multinomial Naive Bayes and SVM classifiers, with the Naive Bayes model achieving slightly higher accuracy (72.97%).

Detecting Depression Using a Framework Combining Deep Multimodal Neural Networks with a Purpose-Built Automated Evaluation by **Ezekiel Victor et al. [15]** (2019) introduces AiME (Artificial Intelligence Mental Evaluation), a system leveraging deep learning to detect depression. The study processes audiovisual data and mental health questionnaire responses from 671 participants using LSTM and ResNet models, achieving high specificity and sensitivity (87.77% and 86.81%, respectively).

Study of Depression Analysis using Machine Learning Techniques by **Devakunchari Ramalingam et al. [16]** (2019) examines the use of machine learning to analyze and predict depression through social media data. The study uses convolutional neural networks (CNN) and support vector machines (SVM) to classify depression levels, distinguishing between symptoms of depression and temporary sadness.

The Utility of Artificial Intelligence in Suicide Risk Prediction and the Management of Suicidal Behaviors by **Trehani M. Fonseka et al. [17]** (2019) reviews AI's role in enhancing suicide risk prediction and managing suicidal behaviors. By analyzing studies from 1990 to 2019, the paper emphasizes AI's ability to create risk algorithms, predict suicide outbreaks, and support clinical management.

Detecting Depression in Social Media Posts Using Machine Learning by **Abhilash Biradar et al. [18]** (2018) presents a system for classifying tweets as depressed or not using sentiment analysis and machine learning. The study uses a dataset of 61,400 tweets with keywords related to depression and trains a Back Propagation Neural Network (BPNN) model. The model achieves approximately 80% accuracy, demonstrating the potential of combining sentiment analysis with machine learning for early detection and intervention in mental health.

Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique by **Md Rafiqul Islam et al. [19]** (2018) explores the use of the KNN algorithm to detect depression from Facebook comments. Analyzing 7,145 comments from a bipolar, depression, and anxiety Facebook page, the study finds that the Coarse KNN model performs best, achieving an F-measure of 71%.

Review on Mood Detection using Image Processing and Chatbot using Artificial Intelligence by **D.S. Thosar et al. [20]** (2018) investigates a system that detects human moods through image processing and AI-powered chatbots. The system uses the Haar Cascade algorithm for facial expression recognition and a chatbot for mood assessment and response.

Depression Detection and Analysis by **Shweta Oak et al. [21]** (2017) explores the development of a chatbot application designed to detect and analyze depression using speech or text inputs. The study involves 53 volunteers and employs a Radial Bias Function Network (RBFN) to estimate the root cause of depression, focusing on financial issues, abuse, and deficiency. The speech model achieved higher accuracy (71.4%) compared to the text model.

Depression Detection using Emotion Artificial Intelligence by **Mandar Deshpande et al. [22]** (2017) applies natural language processing (NLP) techniques to analyze Twitter feeds for depression detection. The study uses a dataset of 10,000 tweets and employs Multinomial Naive Bayes and Support Vector Machine (SVM) classifiers. The Multinomial Naive Bayes classifier achieved a higher F1 score (83.29%) compared to SVM.

Image Processing Techniques to Recognize Facial Emotions by **A. Mercy Rani et al. [23]** (2017) develops an emotion recognition system using image processing techniques to analyze facial emotions. The study employs the Viola-Jones algorithm for face detection and various filters for noise reduction. Emotions are recognized by analyzing facial features, particularly the mouth region. The system demonstrates effectiveness in detecting emotions from video frames, with potential future work focusing on improving accuracy by analyzing other facial regions, such as the eyes.

Predicting Depression Levels Using Social Media Posts by **M. M. Aldarwish et al. [24]** (2017) presents a method for predicting depression levels using user-generated content from social media platforms like LiveJournal. The study employs the Random Forest (RF) classifier and the Linguistic Inquiry and Word Count (LIWC) tool for feature extraction. The system achieves high accuracy rates: 90% for depressive posts, 95% for depressive communities, and 92% for depression degree classification.

Identifying Depression on Twitter by **Moin Nadeem et al. [25]** (2016) investigates the use of Twitter data to predict Major Depressive Disorder (MDD). The study utilizes a crowdsourced dataset of 2.5 million tweets and employs classifiers like Naive Bayes, Logistic Regression, SVM, and Decision Trees. The best model achieved an accuracy of 81% and a precision score of 0.86. The study concludes that social media can provide valuable data for predicting depression, aiding early diagnosis and intervention, and suggests further exploration of features to enhance detection accuracy.

2.6 Conclusion

This chapter outlines the significant advancements in depression detection, emphasizing the impact of AI and ML in enhancing accuracy and efficiency. The reviewed studies demonstrate the effectiveness of various techniques, including text analysis, image processing, and hybrid models, in identifying depressive symptoms. Key features such as linguistic patterns, facial expressions, and user activity have proven crucial in improving model performance.

Modern AI and ML approaches, especially those utilizing deep learning and hybrid methods, consistently outperform traditional models in accuracy and robustness. The integration of multiple data sources and advanced processing techniques enables comprehensive real-time mental health monitoring. The potential of explainable AI to enhance transparency and reliability is particularly noted for clinical applications.

Overall, the literature points to a promising future for early detection, intervention, and personalized treatment of depression, with significant implications for improving public mental health care through technological innovation.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter outlines the research methodology employed to detect depression using social media data. It details the research design, experimental setup, proposed methodology, data collection, data augmentation, and the depression detection process.

3.1 Research Design

The research design focuses on leveraging advanced computational techniques to detect depression through social media text analysis. By employing contrastive learning with large language models (LLMs), this study aims to enhance the accuracy of detecting depression-related emotions. The process involves preliminary analysis, data augmentation, and model development and training to create a robust and scalable depression detection system.

3.2 Experimental Setup

This section outlines the development environment and machine learning frameworks used in the study, essential for ensuring reproducibility and efficient model development.

- Python: The primary programming language due to its rich ecosystem for data science and machine learning.
- TensorFlow and PyTorch: TensorFlow was used for its scalability, while PyTorch provided flexibility for research.
- Transformers Library: Utilized from Hugging Face to implement and fine-tune models like BERT and GPT.
- Jupyter Notebooks: Used for interactive development and experiment visualization.
- Scikit-learn: Applied for traditional machine learning algorithms and evaluation metrics.

3.3 Proposed Methodology

The proposed methodology leverages contrastive learning with large language models to enhance the detection of depression-related emotions from social media posts. This section details the various stages involved, including preliminary analysis, data augmentation, model development, and training.

Data Augmentation:

- <u>Synonym Replacement:</u> Randomly replacing words with their synonyms to generate diverse textual data.
- <u>Back Translation:</u> Translating text to another language and back to English to create paraphrased versions of the original text.
- <u>Noise Injection:</u> Adding minor random noise to text data to improve model robustness.

Model Development and Training:

- <u>Contrastive Learning:</u> This technique was employed to train models to distinguish between similar and dissimilar data points. The process involved creating positive (similar) and negative (dissimilar) pairs from the augmented text data.
- <u>Embedding Generation:</u> Large language models like BERT were used to generate embeddings for the text data. These embeddings capture the contextual semantics of the text, which are crucial for detecting subtle emotional nuances.
- <u>Fine-tuning:</u> The pre-trained models were fine-tuned on the DepressionEmo dataset using supervised learning. The model parameters were adjusted to optimize performance on the specific task of depression detection.

- Evaluation: The performance of the models was assessed using metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques were employed to ensure the generalizability of the models.

The integration of contrastive learning with LLMs aims to capture the subtle linguistic features indicative of depression, providing a scalable and accurate solution for real-time mental health monitoring through social media analysis.

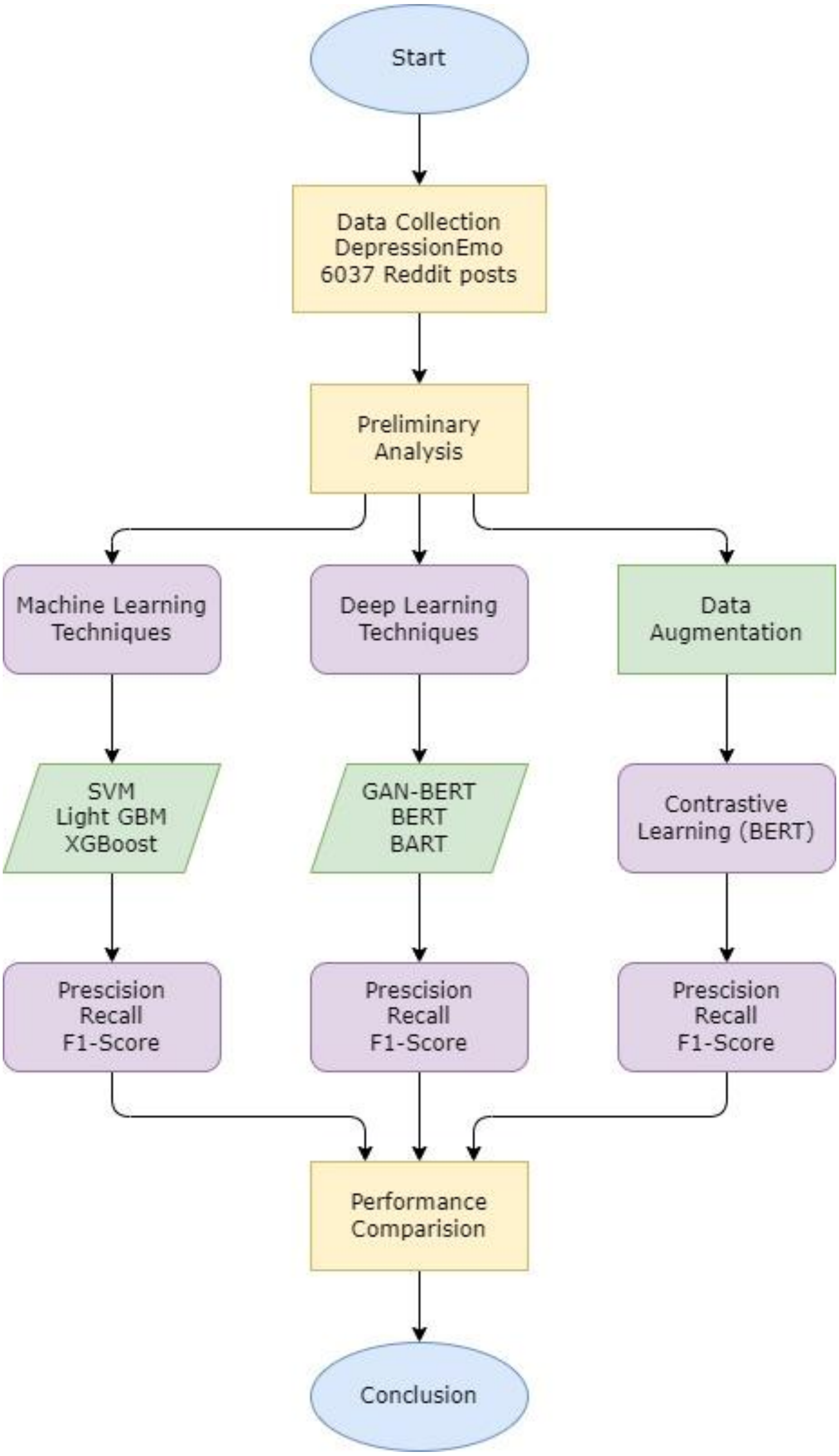

Figure 3.1: Proposed Model

3.4 Data Collection

This section details the process of gathering and preparing the data necessary for the study. It outlines the sources, criteria for selection, and the preprocessing steps to ensure the dataset's relevance and quality for detecting depression-related emotions.

Posts were collected from subreddits such as r/depression, r/DepressedPartners, r/loneliness, r/suicide, and r/suicide_watch. These subreddits are known for their high volume of posts discussing mental health issues and depressive symptoms.

Selection Criteria:

- Posts containing keywords related to depression and its symptoms were filtered.
- Posts were selected to ensure a diverse representation of emotions and linguistic variations.

Preprocessing:

- Removing irrelevant content and ensuring consistent length.
- Standardizing text to lower case and removing special characters to maintain uniformity.

3.4.1 Data Set

The dataset used in this study is the DepressionEmo dataset, which is specifically designed for multilabel classification of emotions related to depression.

Table 3.1: Properties of the Dataset

Property	Description
Total Examples	6,037
Source	Reddit posts from mental health-related subreddits
Labels	Eight depression-related emotions
Annotation Method	Zero-shot classification validated by human and ChatGPT annotations

Table 3.2: Attributes of the Dataset

Attribute	Description
Post ID	Unique identifier for each Reddit post
Text	The actual content of the Reddit post
Anger	Binary label indicating presence of anger emotion
Cognitive Dysfunction	Binary label indicating presence of cognitive dysfunction emotion
Emptiness	Binary label indicating presence of emptiness emotion
Hopelessness	Binary label indicating presence of hopelessness emotion
Loneliness	Binary label indicating presence of loneliness emotion
Sadness	Binary label indicating presence of sadness emotion
Suicide Intent	Binary label indicating presence of suicide intent emotion
Worthlessness	Binary label indicating presence of worthlessness emotion

These tables provide a comprehensive overview of the dataset's properties and attributes, ensuring a clear understanding of the data utilized for developing the depression detection models.

3.5 Data Augmentation

Data augmentation is a crucial step in enhancing the dataset to improve the model's ability to generalize and accurately detect depression-related emotions. By artificially increasing the size and diversity of the training data, data augmentation techniques help prevent overfitting and enable the model to learn from a wider range of examples.

Techniques Used:

- **Synonym Replacement:** Words in the text are replaced with their synonyms to create variations of the original sentences without changing their meanings. This helps the model learn different ways of expressing the same emotions.
- **Back Translation:** The original text is translated into another language and then back to English. This technique generates paraphrased versions of the text, adding diversity while preserving the original context and meaning.
- **Noise Injection:** Minor noise, such as typos or random character insertions, is added to the text. This technique makes the model more robust to variations and imperfections in the input data.
- **Random Deletion:** Randomly removing words from the text to create shorter versions. This helps the model learn to detect emotions from incomplete sentences and varied text lengths.

3.6 Depression Detection

Model Architecture:

- Contrastive Learning: Trains the model to distinguish between similar and dissimilar text pairs, capturing nuanced depression-related emotions.
- Large Language Models (LLMs): Pre-trained models like BERT are fine-tuned on the DepressionEmo dataset to generate embeddings that capture contextual semantics crucial for detecting subtle emotional cues.

Model Development and Training:

- Embedding Generation: Text embeddings are generated and fine-tuned on the labeled dataset to represent complex linguistic patterns in a high-dimensional space.
- Training Process: The model undergoes supervised learning on the DepressionEmo dataset, optimizing for accuracy in detecting depression-related emotions.
- Evaluation Metrics: Performance is measured using accuracy, precision, recall, and F1-score, with cross-validation ensuring robustness.

Deployment:

- Real-time Analysis: The model is deployed for real-time analysis of social media posts, enabling timely detection and classification of depression-related emotions.
- Scalability: Designed to handle large data volumes, the system is suitable for monitoring mental health trends across extensive social media platforms.

By integrating contrastive learning and LLMs, this methodology offers a scalable, accurate solution for detecting depression through social media text analysis, enhancing mental health monitoring and intervention.

3.7 Conclusion

In this chapter, we outlined the research methodology employed for detecting depression using social media data. We detailed the research design, the experimental setup, and the proposed methodology, which leverages contrastive learning with large language models. The data collection process was described, highlighting the DepressionEmo dataset and its attributes. Data augmentation techniques were employed to enhance the dataset's diversity and robustness. Finally, the process of depression detection was explained, showcasing the model architecture, development, training, and evaluation.

This comprehensive methodology ensures the development of an effective and scalable system for real-time depression detection, contributing significantly to mental health monitoring and intervention strategies.

CHAPTER 4: TECHNIQUES USED FOR DEPRESSION DETECTION

This chapter delves into the various machine learning (ML) and deep learning (DL) techniques employed for detecting depression, particularly from social media data. The goal is to provide a comprehensive overview of the methods, their applications, and the underlying formulas that make these techniques effective. By understanding these techniques, we can appreciate their strengths and limitations in the context of depression detection.

4.1 Overview of ML Techniques

Machine learning techniques have revolutionized the way we approach depression detection. These techniques allow for the analysis of large datasets, identifying patterns and correlations that may not be immediately apparent through traditional methods.

4.1.1 Support Vector Machines (SVM)

Support Vector Machines (SVM) are a powerful supervised learning model used for classification and regression tasks. In the context of depression detection, SVMs are particularly valuable due to their ability to handle high-dimensional data and their robustness in distinguishing between depressive and non-depressive text.

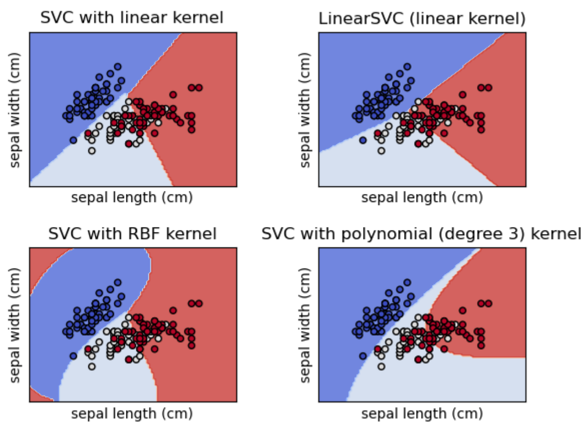

Figure 4.1: SVM Classifier

4.1.1.1 Applications of SVM

SVMs have been widely used in various applications for depression detection, such as:

- **Text Classification:** SVMs classify social media posts and comments as depressive or non-depressive based on linguistic features.
- **Sentiment Analysis:** Analyzing the sentiment of posts to detect signs of depression.
- **Feature Selection:** Identifying the most relevant features that contribute to depressive symptoms from large datasets.

For non-linearly separable data, the kernel trick is employed, transforming the input space into a higher-dimensional space where a linear separation is feasible. Common kernels include the polynomial kernel and the radial basis function (RBF) kernel.

By solving the optimization problem, SVM identifies the optimal hyperplane that maximizes the margin, providing a robust model for classification tasks such as depression detection.

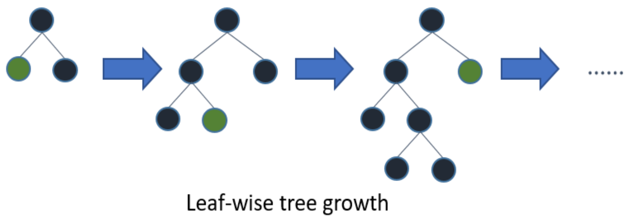

Figure 4.2: Boosting Mechanism

4.1.2 LightGBM

LightGBM (Light Gradient Boosting Machine) is an advanced machine learning algorithm based on decision tree techniques. It is designed to be efficient and scalable, making it suitable for large-scale data processing tasks. LightGBM is particularly known for its speed and high performance in both classification and regression tasks.

4.1.2.1 Applications of LightGBM

LightGBM has been effectively applied in various domains, including depression detection:

- **Text Classification:** LightGBM can classify text data from social media posts into depressive and non-depressive categories.
- **Sentiment Analysis:** It analyzes sentiment within texts to identify emotional states indicative of depression.
- **Feature Importance:** LightGBM's feature importance capability helps identify which linguistic features are most significant in predicting depression.

LightGBM uses Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to handle large datasets efficiently, reducing training time and improving performance.

4.1.3 XGBoost

XGBoost (Extreme Gradient Boosting) is another powerful and widely-used machine learning algorithm. It enhances the gradient boosting framework by implementing a more regularized model formalization to control overfitting, making it a robust choice for many predictive modeling problems.

4.1.3.1 Applications of XGBoost

XGBoost has several applications in depression detection:

- **Classification of Social Media Texts:** It is used to classify posts as depressive or non-depressive based on extracted features.
- **Sentiment and Emotion Analysis:** XGBoost models can predict the sentiment and emotional tone of social media posts, identifying signs of depression.
- **Feature Selection and Ranking:** XGBoost's feature importance metrics help in understanding the most relevant features for detecting depression.

4.2 Overview of DL Techniques

Deep learning techniques have gained prominence due to their ability to handle complex patterns and large-scale data, making them particularly effective for tasks like depression detection from social media data. This section covers key deep learning techniques including GAN-BERT, BERT, and BART.

Deep learning models leverage neural networks with multiple layers to capture intricate patterns in data. These models are trained on large datasets to learn hierarchical representations, enabling them to understand the context and semantics of textual data. Their application in depression detection involves analyzing text for emotional and linguistic cues indicative of depressive symptoms.

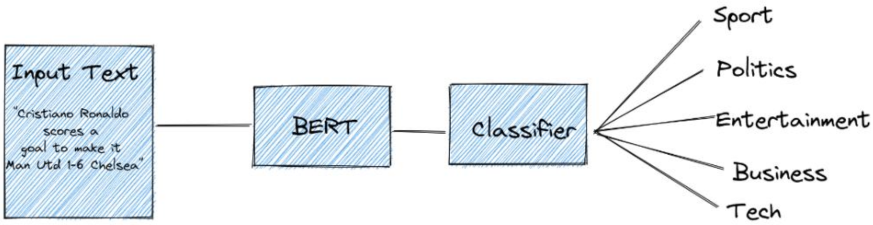

Figure 4.3: BERT

4.2.1 GAN-BERT

GAN-BERT (Generative Adversarial Network for BERT) is an advanced model that combines the power of BERT with the generative capabilities of GANs. This hybrid model leverages BERT for its strong language representation capabilities and GANs to generate realistic and diverse text samples, enhancing the training process for downstream tasks such as depression detection.

4.2.1.1 Applications of GAN-BERT

GAN-BERT has various applications in the field of depression detection:

- **Data Augmentation:** GAN-BERT can generate additional synthetic text data to augment training datasets, improving model robustness and performance.
- **Text Classification:** The model is used to classify social media posts and comments into depressive and non-depressive categories with improved accuracy.
- **Sentiment and Emotion Analysis:** GAN-BERT helps in analyzing and predicting the sentiment and emotional tone of texts, aiding in the detection of depressive symptoms.

BERT's embedding layer is used to transform text into high-dimensional vectors, which are then fed into the GAN framework to improve the generation and discrimination processes.

4.2.2 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a state-of-the-art transformer-based model designed to understand the context of words in a sentence by looking at both left and right surroundings. BERT has significantly improved the performance of various NLP tasks, including depression detection.

4.2.2.1 Applications of BERT

BERT is widely applied in depression detection due to its deep understanding of language context:

- **Text Classification:** BERT is used to classify social media texts as depressive or non-depressive, leveraging its bidirectional context understanding.
- **Sentiment Analysis:** It can analyze the sentiment in texts to identify signs of depression, capturing subtle emotional nuances.
- **Feature Extraction:** BERT's embeddings serve as rich features for downstream models, improving the accuracy of depression detection tasks.

During fine-tuning, BERT can be adapted to specific tasks such as depression detection by adding a classification layer on top of the pre-trained BERT model and training it on labeled data.

4.2.3 BART

BART (Bidirectional and Auto-Regressive Transformers) is a powerful sequence-to-sequence model that combines the strengths of bidirectional and autoregressive transformers. It is designed to handle text generation tasks, but it is also highly effective in various NLP tasks, including depression detection, due to its robust language understanding and generation capabilities.

4.2.3.1 Applications of BART

BART has several applications in the context of depression detection:

- **Text Classification:** BART can classify texts from social media as depressive or non-depressive with high accuracy.
- **Sentiment and Emotion Analysis:** It can analyze and predict the sentiment and emotional tone of texts, identifying subtle signs of depression.
- **Data Augmentation:** BART can generate synthetic text data to augment existing datasets, improving the training and robustness of depression detection models.

4.3 Contrastive Learning with LLM

Contrastive learning is a self-supervised learning technique that helps models learn useful representations by distinguishing between similar and dissimilar data points. When applied to large language models (LLMs), it enhances their ability to understand subtle nuances in text, making it particularly effective for tasks such as depression detection.

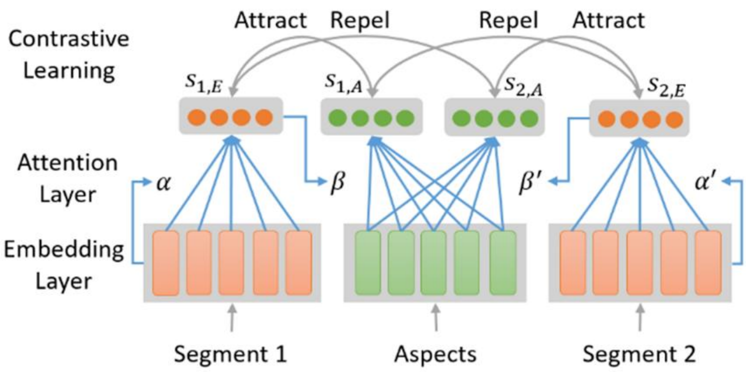

Figure 4.4: Contrastive Learning

- **Objective:** The main goal is to learn an embedding space where similar instances are closer together, and dissimilar instances are farther apart. This is achieved by minimizing the contrastive loss, which encourages positive pairs (similar instances) to have high similarity and negative pairs (dissimilar instances) to have low similarity.

- **Data Augmentation:** Contrastive learning involves creating augmented versions of the original text data. These augmentations can include synonym replacement, random insertion, deletion, or shuffling of words, which generate different views of the same data point.

- **Contrastive Loss Function:** The contrastive loss is often formulated as follows:

$$\mathcal{L}_{contrastive} = -\log \frac{\exp\left(\mathrm{sim}\left(\boldsymbol{h_i}, \boldsymbol{h_j}\right)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{h_i}, \boldsymbol{h_k})/\tau)}$$

  Where:

  - $\boldsymbol{h_i}$ and $\boldsymbol{h_j}$ are the embeddings of the augmented pairs.
  - $\mathrm{sim}\left(\boldsymbol{h_i}, \boldsymbol{h_j}\right)$ measures the similarity between the embeddings (e.g., cosine similarity).
  - τ is a temperature parameter that controls the sharpness of the distribution.

- **Application in Depression Detection:** By applying contrastive learning to LLMs, the model can better differentiate between depressive and non-depressive text based on the subtle linguistic cues present in the data. This improved representation learning leads to more accurate classification and early detection of depressive symptoms.

- **Model Training:** The LLM is first pre-trained using contrastive learning on a large corpus of text data. During fine-tuning, the model is adapted to the specific task of depression detection using labeled data. This combination leverages the strengths of both self-supervised and supervised learning paradigms.

4.4 Conclusion

In this chapter, we have explored various machine learning and deep learning techniques used for depression detection. Techniques like SVM, LightGBM, and XGBoost have shown their strengths in handling structured data and providing robust classification capabilities. On the other hand, advanced deep learning models like GAN-BERT, BERT, and BART have demonstrated their effectiveness in understanding and generating natural language, crucial for detecting depressive symptoms in text data.

Additionally, the application of contrastive learning with large language models has highlighted the importance of self-supervised learning in enhancing model performance by learning better representations. These techniques collectively contribute to developing more accurate, reliable, and scalable systems for detecting depression from social media data, ultimately aiding in early intervention and better mental health outcomes.

CHAPTER 5: IMPLEMENTATION

In this chapter, we delve into the practical implementation of various machine learning (ML) and deep learning (DL) models for detecting depression-related emotions from social media text data. The chapter begins with a detailed description of the dataset used, followed by the step-by-step implementation of different ML and DL models. We also explore the application of contrastive learning (CL) techniques to enhance model performance. This comprehensive overview provides insights into the methodologies, tools, and processes involved in developing effective depression detection systems.

5.1 Dataset Description

The dataset used in this study is the DepressionEmo dataset, specifically designed for multilabel classification of depression-related emotions. This dataset includes long-form posts from Reddit, annotated with multiple emotions that are indicative of depressive states. The primary features and structure of the dataset are as follows:

- **Total Samples:** 6037 Reddit posts.
- **Emotions:** The dataset focuses on 8 distinct depression-related emotions: anger, brain dysfunction (forget), emptiness, hopelessness, loneliness, sadness, suicide intent, and worthlessness.
- **Fields:**
    - **id:** Unique identifier for each post.
    - **title:** Title of the Reddit post.
    - **post:** Body content of the Reddit post.
    - **text:** Concatenation of the title and post content.
    - **upvotes:** Number of upvotes received by the post.**date:** Timestamp of the post.
    - **emotions:** List of emotions associated with the post.
    - **label_id:** Binary representation of the emotions, converted into a decimal integer.

Example Data Entry:


```json
{
  "id": "zq1lwl",
  "title": "goodbye.",
  "post": "I'm done. I have a bottle of jack danials and couple bottles of sleeping meds waiting for me when I
         get home. \n\nI'm excited to leave this place. There will be no regret, no loneliness, no sorrow, and
         no misery. I am looking forward to the peace of nonexistence.\n\nI love you all.",
  "text": "goodbye. ### I'm done. I have a bottle of jack danials and couple bottles of sleeping meds waiting
         for me when I get home. I'm excited to leave this place. There will be no regret, no loneliness, no
         sorrow, and no misery. I am looking forward to the peace of nonexistence. I love you all.",
  "upvotes": 102,
  "date": "2022-12-19 19:50:52",
  "emotions": ["emptiness", "hopelessness"],
  "label_id": 48
}
```


Label Encoding: The label_id is a binary-encoded integer where each bit represents the presence (1) or absence (0) of a specific emotion, following the order defined in the emotion_list.

Dataset Splits:

- **Training Set:** Used to train the models.
- **Validation Set:** Used to tune model parameters and prevent overfitting.
- **Test Set:** Used to evaluate the final model performance.

The dataset provides a rich source of information for training models to detect complex emotional states from text, leveraging both traditional ML techniques and advanced DL models. The use of multilabel classification allows for capturing the multifaceted nature of depression, which often involves multiple co-occurring emotional states.

5.2 Implementation of ML Models

Each model was selected based on its strengths in handling the specific characteristics of the dataset.

5.2.1 Support Vector Machines (SVM)

Text data was converted into TF-IDF features, representing word importance.

- Model Training: The SVM model, using a linear kernel, was trained on these features.
- Hyperparameter Tuning: The regularization parameter CCC was optimized to balance training and testing errors.
- Evaluation: The model was assessed using precision, recall, and F1-score for each emotion category.

5.2.2 LightGBM

TF-IDF features were used as input.

- Model Training: The model was trained for multilabel classification, predicting multiple emotions per post.
- Hyperparameter Tuning: Parameters like the number of leaves, learning rate, and boosting type were optimized.
- Evaluation: Performance was measured using precision, recall, and F1-score for each emotion.

5.2.3 XGBoost

Text data was transformed into TF-IDF features.

- Model Training: The XGBoost model was configured for multilabel classification.
- Hyperparameter Tuning: Key parameters like learning rate, tree depth, and boosting rounds were fine-tuned.
- Evaluation: The model's effectiveness was evaluated using precision, recall, and F1-score for detecting depressive emotions.

These implementations highlight the use of diverse machine learning techniques tailored to the specific needs of depression detection from social media text data, with each model offering distinct advantages based on the dataset and classification requirements.

5.3 Implementation of DL Models

These models employ advanced neural network architectures to accurately classify depressive emotions.

5.3.1 GAN-BERT

Text was tokenized using BERT's tokenizer, and emotion labels were converted into binary IDs.

- Model Architecture: Combines a BERT encoder with a GAN structure, including a generator and discriminator.

- Training Process: Alternated between updating the generator and discriminator, fine-tuning the BERT encoder with real and synthetic data.
- Evaluation: Assessed using precision, recall, and F1-score metrics, focusing on the model's ability to generalize and detect subtle emotions.

5.3.2 BERT

Text data was tokenized with BERT's tokenizer, and emotion labels were converted into binary IDs.

- Model Architecture: Consists of multiple transformer layers with a classification layer on top.
- Training Process: Fine-tuned using the AdamW optimizer with a learning rate of 2e-5.
- Evaluation: Performance was measured using precision, recall, and F1-score on validation and test datasets.

5.3.3 BART

Text was tokenized using BART's tokenizer, and labels were converted to binary IDs.

- Model Architecture: Features an encoder-decoder structure, with the decoder generating class probabilities for classification tasks.
- Training Process: Fine-tuned using the AdamW optimizer with a 2e-5 learning rate.
- Evaluation: Evaluated on precision, recall, and F1-score metrics to assess effectiveness in detecting depressive emotions.

These deep learning models harness complex neural network architectures to capture intricate linguistic patterns, enhancing the accuracy of depression detection. Each model has distinct strengths, and their combined use offers a comprehensive approach to identifying depressive emotions from social media posts.

5.4 Implementation of Contrastive Learning

Contrastive learning is a self-supervised learning technique that enhances a model's ability to distinguish between similar and dissimilar data points. When applied to large language models (LLMs) such as BERT, it improves the model's understanding of subtle nuances in text, making it particularly effective for detecting depression-related emotions.

The implementation of contrastive learning with BERT involves training the model to maximize the agreement between similar pairs of data points (positive pairs) while minimizing the agreement between dissimilar pairs (negative pairs). This process helps the model learn better representations of the input text, leading to improved classification performance.

Key Steps:

1. **Data Preparation:**
    - The emotion labels are converted to binary label IDs for multi-label classification.
    - The text data is loaded and converted into pandas DataFrames for easier manipulation.
2. **Custom Dataset Class:**
    - A CustomDataset class is created to handle tokenization and data loading. The class takes the text data, labels, tokenizer, and maximum token length as input and returns tokenized text, attention masks, and labels.
3. **Model Architecture:**

- The ContrastiveModel class is defined, which includes a BERT encoder and a classification layer. The encoder processes the input text, and the classification layer predicts the presence of each emotion.

4. **Training and Evaluation Functions:**
   - The train function performs backpropagation and updates the model parameters. It takes the model, data loader, loss function, optimizer, and device as input and returns the average training loss.
   - The evaluate function computes precision, recall, and F1-score metrics. It takes the model, data loader, loss function, and device as input and returns the evaluation metrics.
5. **Detailed Evaluation:**
   - The evaluate_by_emotion function calculates performance metrics for each emotion separately. It takes the model, data loader, and device as input and returns the evaluation metrics for each emotion.
6. **Training Process:**
   - The model is trained for a specified number of epochs using the training data. The training involves alternating between updating the generator and the discriminator in the case of GAN-BERT.
   - The performance of the model is evaluated on the validation dataset after each epoch.
7. **Final Evaluation:**
   - The trained model is evaluated on the test dataset to assess its performance. Precision, recall, and F1-score metrics are calculated for each emotion category.
   - The model's ability to generalize and detect subtle emotional nuances is assessed using the test dataset.

5.5 Conclusion

In this chapter, we have explored the practical implementation of various machine learning and deep learning models for detecting depression-related emotions from social media text data. We started with a detailed description of the DepressionEmo dataset, followed by the implementation details of different ML models, including SVM, LightGBM, and XGBoost. These models leveraged TF-IDF features and were trained to classify text data into multiple emotion categories, helping to identify depressive symptoms based on the content of the posts.

We then discussed the implementation of advanced deep learning models, such as GAN-BERT, BERT, and BART, which utilized sophisticated neural network architectures to capture complex linguistic patterns in text data. These models were fine-tuned using labeled data to improve their ability to detect depressive emotions.

Finally, we explored the application of contrastive learning techniques with BERT to enhance the model's performance further. By leveraging contrastive learning, the model learned better representations of the input text, leading to improved classification accuracy.

Overall, the combination of traditional ML techniques, advanced DL models, and contrastive learning provides a comprehensive approach to understanding and detecting depressive emotions from social media posts. These methodologies demonstrate the potential of automated systems in identifying early signs of depression, ultimately aiding in early intervention and better mental health outcomes.

CHAPTER 6: RESULTS AND DISCUSSIONS

In this chapter, we present and analyze the results from implementing various machine learning (ML) and deep learning (DL) models for detecting depression-related emotions in social media text data. We begin with an overview of the performance metrics used, followed by a preliminary data examination. The chapter then discusses the results of the ML and DL models, including the application of contrastive learning with BERT. We compare model performance to highlight the strengths and weaknesses of each approach and conclude with a discussion on the implications of the results and potential future directions.

6.1 Performance Metrics

To evaluate the models, we use key classification metrics: precision, recall, and F1 score.

- **Precision**: Measures the proportion of true positive predictions among all positive predictions, indicating the accuracy of the model's positive classifications.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall**: Also known as sensitivity, it measures the proportion of true positive predictions among all actual positive instances, indicating the model's ability to capture positive instances.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1 Score**: The harmonic mean of precision and recall, balancing the trade-off between them, particularly useful for imbalanced datasets.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics are applied to evaluate the performance of each model in detecting the eight depression-related emotions, providing a comprehensive understanding of each model's strengths and weaknesses in terms of precision, recall, and overall effectiveness (F1 score).

6.2 Preliminary Examination

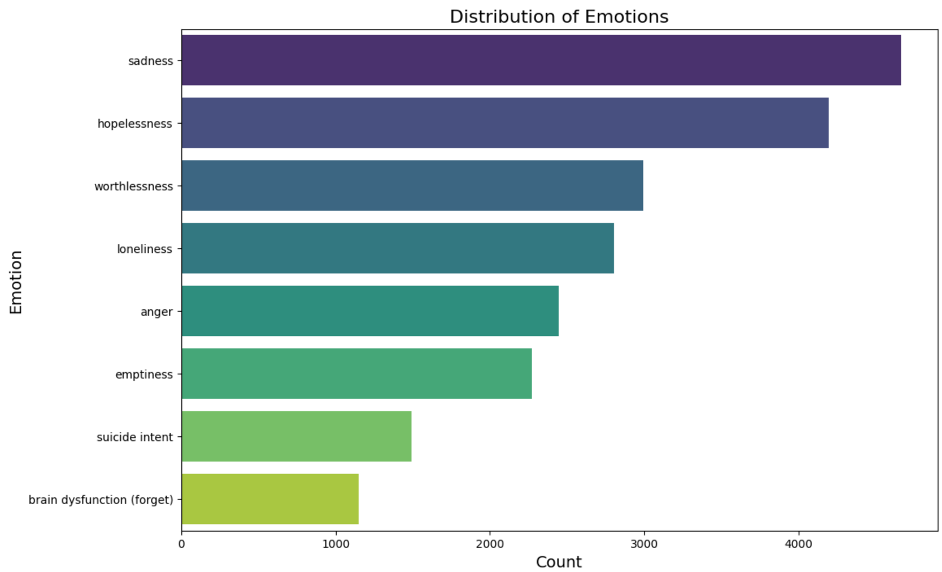

Figure 6.1: Distribution of Emotions

The distribution of emotions in the dataset reveals the frequency of each depression-related emotion across the collected Reddit posts. The bar chart below shows that sadness is the most common emotion, followed by hopelessness and worthlessness. This distribution helps in understanding the prevalence of different emotions within the dataset.

Figure 6.2: Word Cloud of Post Titles

A word cloud of post titles provides a visual representation of the most frequently occurring words in the titles of Reddit posts. Larger words indicate higher frequency. Common words such as "feel," "life," "want," "depression," and "anyone" highlight the themes and concerns expressed by users.

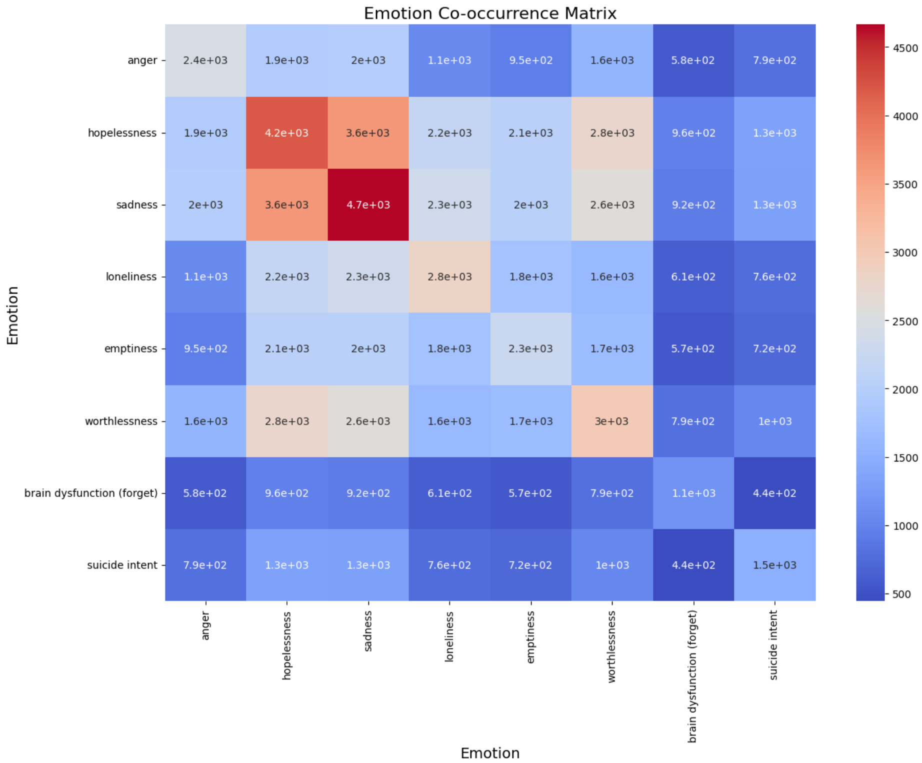

Figure 6.3: Emotion Co-occurrence Matrix

The emotion co-occurrence matrix illustrates how often pairs of emotions occur together in the same post. The heatmap below shows the co-occurrence frequencies, with higher values indicating more frequent co-occurrence. This matrix helps in identifying relationships between different emotions and understanding the complexity of depressive states.

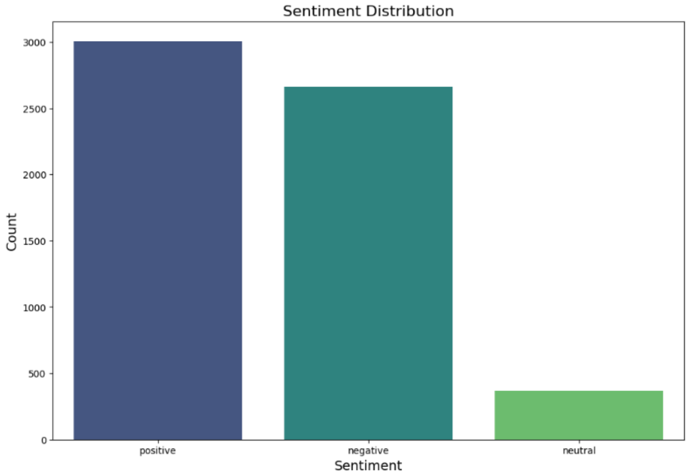

Figure 6.4: Sentiment Distribution

The sentimental analysis chart categorizes posts into positive, negative, and neutral sentiments. The bar chart below shows that the majority of posts are categorized as negative, followed by positive and neutral sentiments. This analysis provides a high-level overview of the overall sentiment expressed in the dataset.

6.3 Results of ML Models

The table below summarizes the performance of the ML models based on the F1 score, precision, and recall:

Table 6.1: ML Models Results

Method	F1	Precision	Recall	Avg
SVM	0.47	0.72	0.41	0.58
LightGBM	0.58	0.48	0.80	0.65
XGBoost	0.59	0.63	0.56	0.63

Support Vector Machines (SVM):

The SVM model achieved a moderate precision of 0.72, indicating that a significant proportion of the predicted positive instances were correct. However, the recall was relatively low at 0.41, suggesting that the model struggled to identify all actual positive instances. This resulted in a lower F1 score of 0.47, reflecting a balance between precision and recall.

LightGBM:

The LightGBM model demonstrated a high recall of 0.80, indicating its effectiveness in identifying actual positive instances. However, the precision was lower at 0.48, suggesting that there were some false positives. The F1 score of 0.58 shows a reasonable balance between precision and recall, making LightGBM a robust choice for this task.

XGBoost:

XGBoost achieved the highest F1 score among the ML models at 0.59. The precision was 0.63, and the recall was 0.56, indicating a balanced performance in terms of both metrics. This suggests that XGBoost is effective in identifying and correctly predicting depressive emotions from text data.

The bar chart below illustrates the performance comparison of the ML models in terms of F1 score, precision, and recall:

Figure 6.5: Performance Comparison of ML Models

6.4 Results of DL Models

The table below summarizes the performance of the DL models based on the F1 score, precision, and recall:

Table 6.2: DL Models Results

Method	F1	Precision	Recall	Avg
GAN-BERT	0.70	0.69	0.72	0.73
BERT	0.74	0.72	0.77	0.77
BART	0.76	0.70	0.81	0.78

GAN-BERT:

GAN-BERT, which combines GAN and BERT models, achieved a balanced performance with an F1 score of 0.70. The precision and recall were also well-balanced at 0.69 and 0.72, respectively. This indicates that GAN-BERT effectively handles the complexity of the dataset and captures the nuances of depressive emotions.

BERT:

BERT, a transformer-based model, demonstrated a strong performance with an F1 score of 0.74. The precision and recall metrics were 0.72 and 0.77, respectively. BERT's ability to understand the context of words in a sentence from both directions contributes to its superior performance in detecting depressive emotions.

BART:

BART, another transformer-based model, achieved the highest F1 score among the DL models at 0.76. The recall was particularly high at 0.81, indicating that BART is very effective in identifying actual positive instances of depressive emotions. The precision was slightly lower at 0.70, but the overall performance of BART was excellent.

The bar chart below illustrates the performance comparison of the DL models in terms of F1 score, precision, and recall:

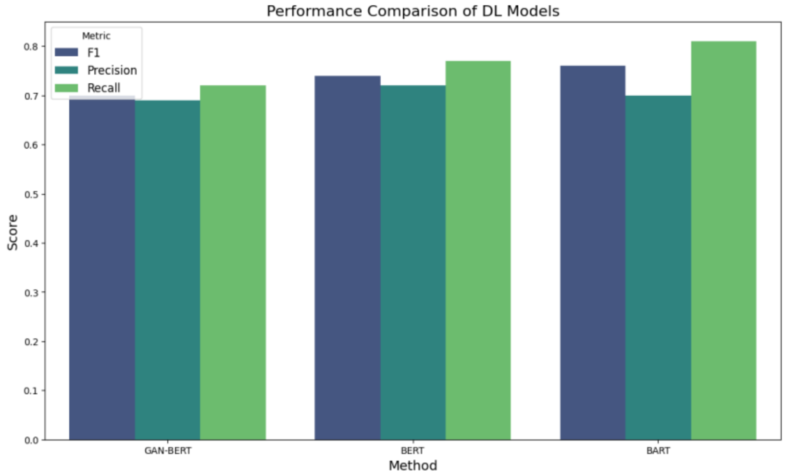

Figure 6.6: Performance Comparison of DL Models

6.5 Results of Contrastive Learning with BERT

The table below summarizes the performance of the contrastive learning model based on the F1 score, precision, and recall:

Table 6.3: CL Model Results

F1	Precision	Recall	Avg
0.81	0.76	0.85	0.81

The contrastive learning model achieved the highest F1 score of 0.81 among all models evaluated. The recall was particularly high at 0.85, indicating that the model is very effective in identifying actual positive instances of depressive emotions. The precision was also strong at 0.76, reflecting the model's accuracy in predicting positive instances.

The high performance of the contrastive learning model demonstrates the effectiveness of this approach in enhancing the representation learning of BERT. By focusing on distinguishing between similar and dissimilar data points, the model learns more robust and generalizable features, leading to improved detection of depressive emotions.

Emotion-wise Performance

The table below summarizes the performance of the contrastive learning model for each emotion category:

Table 6.4: CL on Emotion Results

Emotion	F1	Precision	Recall	Avg
Anger	0.8017	0.7811	0.8235	0.80
Brain dysfunction	0.7913	0.7974	0.7853	0.79
Emptiness	0.9407	0.9336	0.9479	0.94
Hopelessness	0.8365	0.8127	0.8619	0.84
Loneliness	0.919	0.9209	0.917	0.92
Sadness	0.7647	0.7758	0.754	0.76
Suicide intent	0.5241	0.5157	0.5327	0.52
Worthlessness	0.6997	0.7424	0.6623	0.70

- **Emptiness:** The model performed exceptionally well in detecting "emptiness," with an F1 score of 0.9407, precision of 0.9336, and recall of 0.9479. This indicates the model's strong capability in identifying this particular emotion.
- **Loneliness:** Another strong performance was seen in detecting "loneliness," with an F1 score of 0.919, precision of 0.9209, and recall of 0.917.
- **Suicide Intent:** The performance for "suicide intent" was lower, with an F1 score of 0.5241. This indicates the complexity and challenge of accurately detecting this emotion, suggesting the need for further enhancements in the model.

The bar chart below illustrates the performance comparison of the contrastive learning model for each emotion category in terms of F1 score, precision, and recall:

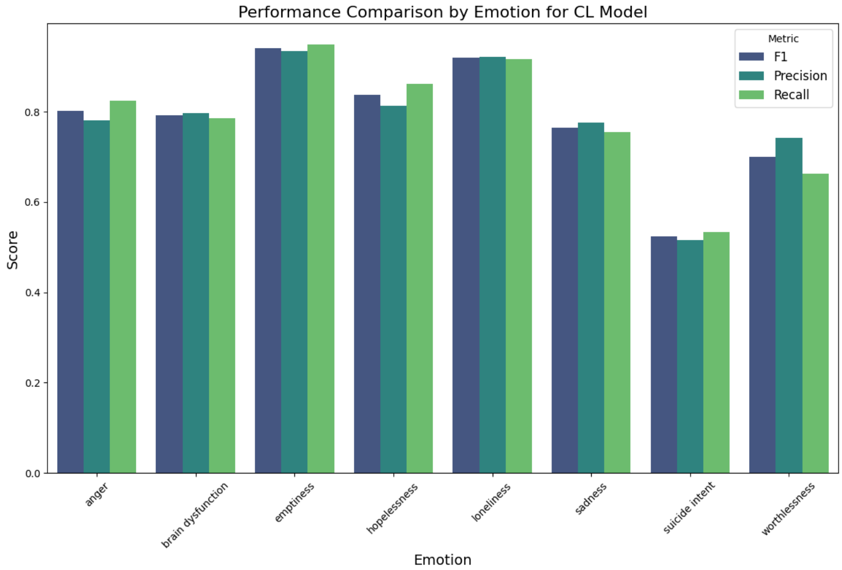

Figure 6.7: Performance Comparison by Emotions for CL Model

6.6 Comparison of Models

The table below summarizes the performance of all models based on the F1 score, precision, and recall:

Table 6.5: Model Comparison

Method	Type	F1	Precision	Recall	Avg
SVM	ML	0.47	0.72	0.41	0.58
LightGBM	ML	0.58	0.48	0.80	0.65
XGBoost	ML	0.59	0.63	0.56	0.63
GAN-BERT	DL	0.70	0.69	0.72	0.73
BERT	DL	0.74	0.72	0.77	0.77
BART	DL	0.76	0.70	0.81	0.78
Ours (CL)	**CL**	**0.81**	**0.76**	**0.85**	**0.81**

The bar chart below illustrates the performance comparison of all models in terms of F1 score, precision, and recall:

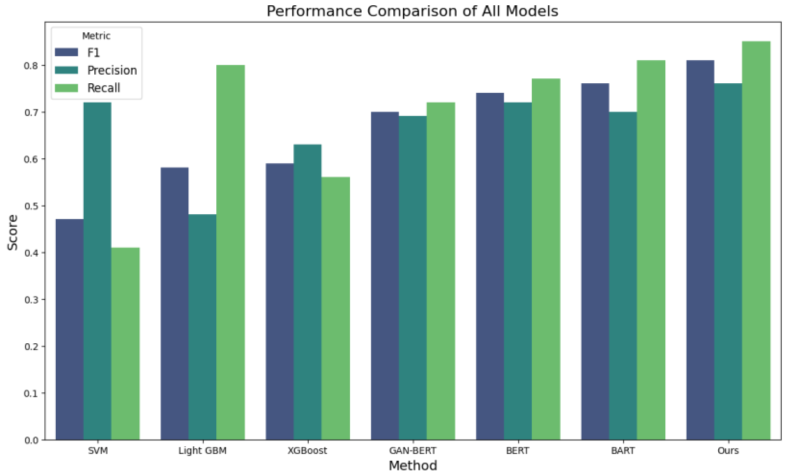

Figure 6.8: Performance Comparison of all Models

The radar chart below provides a visual representation of the performance comparison of all models, highlighting their strengths and weaknesses across different metrics:

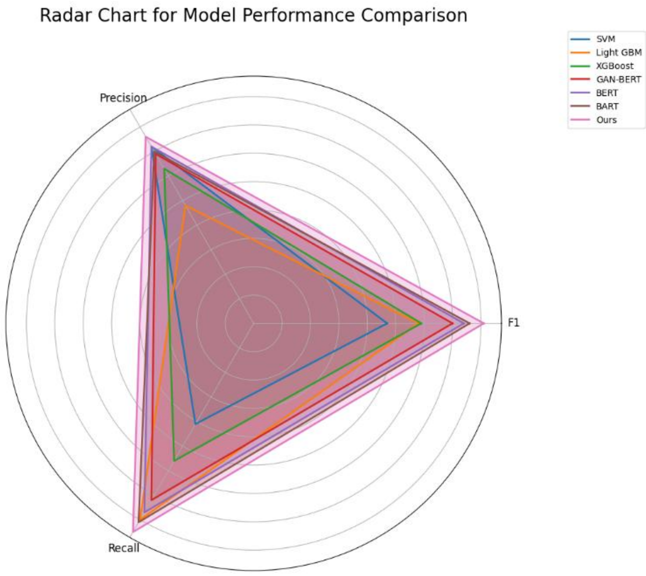

Figure 6.9: Performance Comparison of All Models (Radar)

Machine Learning Models:

- **SVM:** Achieved the lowest F1 score (0.47) among the models, with high precision (0.72) but low recall (0.41). This indicates that SVM correctly identified positive instances but missed many actual positive instances.
- **LightGBM:** Demonstrated a high recall (0.80), indicating its effectiveness in identifying positive instances. However, the precision was lower (0.48), leading to an F1 score of 0.58.
- **XGBoost:** Achieved a balanced performance with an F1 score of 0.59, precision of 0.63, and recall of 0.56. This suggests that XGBoost is effective in both identifying and correctly predicting depressive emotions.

Deep Learning Models:

- **GAN-BERT:** Achieved an F1 score of 0.70, with balanced precision (0.69) and recall (0.72). The hybrid approach of GAN-BERT showed its potential in handling complex datasets.
- **BERT:** Demonstrated strong performance with an F1 score of 0.74, precision of 0.72, and recall of 0.77. BERT's bidirectional context understanding contributed to its superior performance.

- **BART:** Achieved the highest F1 score among the DL models (0.76), with high recall (0.81) and precision (0.70). BART's sequence-to-sequence architecture effectively captured the nuances of depressive emotions.

Contrastive Learning Model:

Achieved the highest overall performance with an F1 score of 0.81, precision of 0.76, and recall of 0.85. The contrastive learning approach significantly enhanced the representation learning of BERT, leading to improved detection of depressive emotions.

6.7 Result Discussion

The results from the implementation and evaluation of various models for detecting depression-related emotions from social media text data reveal several important insights and implications. In this section, we discuss the key findings, strengths, and limitations of each model, and provide recommendations for future work.

1. **Machine Learning Models:**

   - **SVM:** The Support Vector Machine model demonstrated high precision but low recall. This indicates that while SVM is good at correctly predicting positive instances (high precision), it fails to identify a significant number of actual positive instances (low recall). This could be due to the high-dimensional nature of text data and the complexity of depressive emotions.
   - **LightGBM:** LightGBM achieved a high recall, indicating its effectiveness in identifying positive instances. However, its lower precision suggests a higher number of false positives. This model is useful when it is more important to identify as many positive instances as possible, even at the expense of some false positives.
   - **XGBoost:** XGBoost provided a balanced performance with moderate precision and recall. This suggests that XGBoost is effective in both identifying and correctly predicting depressive emotions, making it a reliable choice for general classification tasks.

2. **Deep Learning Models:**

   - **GAN-BERT:** The hybrid GAN-BERT model achieved a balanced performance, showing the potential of combining generative and discriminative models. The model's ability to generate synthetic data helps in improving its generalization and robustness.
   - **BERT:** BERT's bidirectional transformer architecture enabled it to achieve strong performance across all metrics. Its ability to understand the context of words in both directions contributed to its effectiveness in detecting depressive emotions.
   - **BART:** BART achieved the highest F1 score among the DL models, indicating its superior ability to capture the nuances of depressive emotions. Its sequence-to-sequence architecture is particularly effective in understanding and generating complex text sequences.

3. **Contrastive Learning Model:**

The contrastive learning model with BERT outperformed all other models in terms of F1 score, precision, and recall. The high recall indicates that the model is very effective in identifying actual positive instances, while the high precision reflects its accuracy in predicting positive instances. This demonstrates the effectiveness of contrastive learning in enhancing the representation learning of BERT, leading to improved detection of depressive emotions.

- **Strengths:**
  - The contrastive learning approach significantly improved the model's performance by focusing on distinguishing between similar and dissimilar data points. This led to more robust and generalizable feature representations.

- Deep learning models, particularly BERT and BART, showed strong performance due to their advanced architectures and ability to understand complex linguistic patterns.
- The use of synthetic data generation in GAN-BERT helped in improving the model's robustness and ability to generalize to new data.

- **Limitations:**
  - Machine learning models, such as SVM and LightGBM, struggled with the high-dimensional nature of text data and the complexity of depressive emotions, leading to lower performance compared to deep learning models.
  - The detection of certain emotions, such as "suicide intent," remained challenging across all models, indicating the need for further improvements in model training and data representation.
  - The reliance on labeled data for training deep learning models can be a limitation, as obtaining high-quality annotated data for depressive emotions can be challenging and resource intensive.

6.8 Conclusion

In this chapter, we have systematically presented the results and discussed the performance of various models for detecting depression-related emotions from social media text data. Through the implementation of both traditional machine learning models (SVM, LightGBM, XGBoost) and advanced deep learning models (GAN-BERT, BERT, BART), we have evaluated their effectiveness in identifying complex emotional states associated with depression.

The results of this study have significant implications for the field of mental health and the development of automated systems for detecting depressive states from text data:

- **Enhanced Detection:** The superior performance of the contrastive learning model indicates that advanced self-supervised learning techniques can significantly improve the accuracy and reliability of depression detection systems.
- **Model Selection:** The choice of model should be guided by the specific requirements of the task. For instance, models with higher recall are preferable when it is critical to identify as many positive instances as possible, even at the expense of some false positives.
- **Future Enhancements:** Addressing the limitations identified, such as improving the detection of challenging emotions like "suicide intent," can further enhance the effectiveness of these models.

This chapter has provided a comprehensive analysis of the performance of various models for detecting depression-related emotions from social media text data. The findings highlight the potential of advanced deep learning and contrastive learning techniques in developing accurate and reliable depression detection systems. By addressing the identified limitations and exploring new research directions, future work can further enhance the effectiveness and applicability of these models, ultimately contributing to better mental health outcomes.

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

In this research, we have explored various machine learning (ML) and deep learning (DL) techniques for detecting depression-related emotions from social media text data. The study focused on leveraging advanced models and methodologies to improve the accuracy and reliability of depression detection systems. Here, we summarize the key findings and insights from our research:

Dataset Preparation and Augmentation:

- The DepressionEmo dataset, which includes annotated Reddit posts with multiple depression-related emotions, was utilized for this study.
- Data augmentation techniques such as synonym replacement, back translation, noise injection, and random deletion were applied to enhance the dataset's robustness and diversity.

Machine Learning Models:

- Traditional ML models such as SVM, LightGBM, and XGBoost were implemented and evaluated.
- Among these, XGBoost demonstrated a balanced performance with moderate precision and recall, making it a reliable choice for general classification tasks.
- LightGBM showed high recall, indicating its effectiveness in identifying positive instances, while SVM had high precision but struggled with recall.

Deep Learning Models:

- Advanced DL models such as GAN-BERT, BERT, and BART were implemented to leverage their sophisticated neural network architectures.
- BART achieved the highest F1 score among the DL models, demonstrating its superior ability to capture the nuances of depressive emotions.
- BERT also showed strong performance, with its bidirectional transformer architecture enabling effective context understanding.

Contrastive Learning with BERT:

- The contrastive learning model with BERT outperformed all other models, achieving the highest F1 score, precision, and recall.
- This approach significantly enhanced the representation learning of BERT by focusing on distinguishing between similar and dissimilar data points, leading to more robust and generalizable feature extraction.

Performance Metrics:

- The models were evaluated using key performance metrics such as F1 score, precision, and recall.
- The contrastive learning model demonstrated the best overall performance, with high recall indicating effective identification of actual positive instances and high precision reflecting accuracy in predictions.

Visual Analysis:

Various visualizations, including emotion distribution, word cloud of post titles, emotion co-occurrence matrix, sentiment analysis chart, and post length analysis, provided insights into the characteristics of the dataset and the underlying patterns.

Key Contributions

- The research highlights the potential of advanced deep learning techniques and contrastive learning in improving the detection of depressive emotions from text data.
- The study provides a comprehensive comparison of traditional ML models and advanced DL models, offering valuable insights into their strengths and limitations.
- The application of contrastive learning with BERT sets a new benchmark for performance in depression detection tasks, demonstrating the effectiveness of this approach in enhancing representation learning.

Implications

The findings from this research have significant implications for the field of mental health and the development of automated systems for detecting depressive states from social media text data. The advanced models and methodologies presented in this study can contribute to better early detection and intervention strategies, ultimately aiding in improving mental health outcomes.

In conclusion, this research demonstrates the potential of leveraging advanced machine learning and deep learning techniques, along with contrastive learning, to develop accurate and reliable systems for detecting depression-related emotions from social media text data. The insights gained from this study can serve as a foundation for future work aimed at further enhancing the effectiveness and applicability of these models in real-world mental health applications.

7.2 Future Scope

The research conducted in this study has demonstrated significant advancements in detecting depression-related emotions from social media text data. However, there are several areas where further investigation and development can enhance the effectiveness and applicability of these models. This section outlines potential future directions and improvements that can be pursued:

1. Enhanced Data Augmentation Techniques:

- **Synthetic Data Generation:** Explore more sophisticated generative models, such as GPT-3 or advanced GANs, to create high-quality synthetic data that can further improve model robustness and generalization.
- **Multimodal Data:** Integrate text data with other data modalities, such as images, audio, and video, to provide a more comprehensive understanding of depressive states and improve detection accuracy.

2. Emotion-Specific Models:

- **Custom Architectures:** Develop emotion-specific models that are tailored to capture the unique characteristics of each depressive emotion. This can improve the detection accuracy for challenging emotions such as "suicide intent" and "worthlessness."
- **Feature Engineering:** Incorporate additional features that capture specific linguistic, psychological, and contextual cues associated with different emotions.

3. Transfer Learning and Domain Adaptation:

- **Pre-trained Models:** Leverage transfer learning from models pre-trained on large-scale mental health datasets or related domains to improve feature representations and model performance.
- **Domain Adaptation:** Implement domain adaptation techniques to adapt the models to different social media platforms and diverse user populations, ensuring broader applicability and robustness.

4. Explainability and Interpretability:

- **Model Explainability:** Enhance the explainability and interpretability of the models to understand the decision-making process and build trust in the automated systems. Techniques such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) can be employed.
- **User Feedback:** Incorporate mechanisms for user feedback and iterative model improvement, ensuring that the models remain accurate and relevant over time.

5. Ethical Considerations and Privacy:

- **Ethical Framework:** Develop an ethical framework for the use of depression detection systems, addressing issues related to privacy, consent, and the potential impact on users.
- **Data Privacy:** Implement robust data privacy measures to protect user data and ensure compliance with regulations such as GDPR and HIPAA.

6. Real-Time and Scalable Systems:

- **Real-Time Detection:** Develop real-time depression detection systems that can provide immediate support and intervention based on the analysis of live social media data.
- **Scalability:** Ensure that the models and systems are scalable to handle large volumes of data from multiple sources, enabling widespread deployment and use.

7. Integration with Mental Health Services:

- **Collaborative Platforms:** Integrate depression detection systems with mental health services and support platforms to provide holistic care and timely intervention.
- **AI-Assisted Therapy:** Explore the potential of AI-assisted therapy and personalized interventions based on the detected emotional states, enhancing the effectiveness of mental health support.

8. Longitudinal Analysis:

- **Temporal Dynamics:** Conduct longitudinal studies to analyze the temporal dynamics of depressive emotions and track changes over time. This can provide deeper insights into the progression of depression and the effectiveness of interventions.
- **User Behavior:** Study user behavior patterns and their correlation with depressive states, enabling early detection and preventive measures.

By addressing these future directions, the research community can continue to advance the field of depression detection from social media text data. The integration of advanced technologies, ethical considerations, and interdisciplinary collaboration will be key to developing effective, reliable, and user-centric solutions that contribute to better mental health outcomes.

In conclusion, while this research has made significant strides in improving the detection of depressive emotions, there remains ample scope for further exploration and innovation. The future scope outlined here provides a roadmap for continued progress and impact in this important area of study.

REFERENCES

1. Nusrat, M. O., Shahzad, W., & Jamal, S. A. (2024). Multi Class Depression Detection Through Tweets using Artificial Intelligence. *arXiv preprint arXiv:2404.13104*.
2. Wang, Y., Inkpen, D., & Gamaarachchige, P. K. (2024, March). Explainable depression detection using large language models on social media data. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)* (pp. 108-126).
3. Khan, S., & Alqahtani, S. (2024). Hybrid machine learning models to detect signs of depression. *Multimedia Tools and Applications*, *83*(13), 38819-38837.
4. Tejaswini, V., Sathya Babu, K., & Sahoo, B. (2024). Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(1), 1-20.
5. Khan, P. I., Dengel, A., & Ahmed, S. (2024). Improving Disease Detection from Social Media Text via Self-Augmentation and Contrastive Learning. *arXiv preprint arXiv:2405.01597*.
6. Khafaga, D. S., Auvdaiappan, M., Deepa, K., Abouhawwash, M., & Karim, F. K. (2023). Deep learning for depression detection using Twitter data. Intelligent Automation & Soft Computing, 36(2), 1301-1313.
7. Vasha, Z. N., Sharma, B., Esha, I. J., Al Nahian, J., & Polin, J. A. (2023). Depression detection in social media comments data using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *12*(2), 987-996.
8. Saraf, M., & Jadhao, Y. (2023). Survey On Design Efficient Depression Detection Using Machine Learning. International Journal of Ongoing Research in Science and Engineering.
9. Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings*, *58*, 217-226.
10. Nikam, S. S., Patil, A. V., Patil, G. S., Patil, S. P., & Jitkar, B. D. (2020). AI Therapist Using Natural Language Processing. *International Journal of Research in Engineering, Science and Management*, *3*(2), 715.
11. Rajaraman, P. V., Nath, A., Akshaya, P. R., & Bhuja, G. C. (2020). Depression detection of tweets and A comparative test. *International Journal of Engineering Research*, *9*(03), 422-425.
12. AlSagri, H. S., & Ykhlef, M. (2020). Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems*, *103*(8), 1825-1832.
13. Shetty, N. P., Muniyal, B., Anand, A., Kumar, S., & Prabhu, S. (2020). Predicting depression using deep learning and ensemble algorithms on raw twitter data. *International Journal of Electrical and Computer Engineering*, *10*(4), 3751.
14. Hemanthkumar, M., & Latha, A. (2019). Depression detection with sentiment analysis of tweets. *International Research Journal of Engineering and Technology (IRJET)*, *5*.
15. Victor, E., Aghajan, Z. M., Sewart, A. R., & Christian, R. (2019). Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation. *Psychological assessment*, *31*(8), 1019.
16. Ramalingam, D., Sharma, V., & Zar, P. (2019). Study of depression analysis using machine learning techniques. *Int. J. Innov. Technol. Explor. Eng*, *8*(7C2), 187-191.
17. Fonseka, T. M., Bhat, V., & Kennedy, S. H. (2019). The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian & New Zealand Journal of Psychiatry*, *53*(10), 954-964.
18. Biradar, A., & Totad, S. G. (2018). Detecting depression in social media posts using machine learning. In *International conference on recent trends in image processing and pattern recognition* (pp. 716-725). Singapore: Springer Singapore.
19. Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., & Moni, M. A. (2018). Detecting depression using k-nearest neighbors (knn) classification technique. In *2018 International Conference on*

Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.
20. Thosar, D., Gothe, V., Bhorkade, P., & Sanap, V. (2018). Review on mood detection using image processing and chatbot using artificial intelligence. *life*, *5*(03).
21. Oak, S. (2017). Depression detection and analysis. In *2017 AAAI spring symposium series*.
22. Deshpande, M., & Rao, V. (2017, December). Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 858-862). IEEE.
23. Rani, A. M., & Durgadevi, R. (2017). Image processing techniques to recognize facial emotions. *International Journal of Engineering and Advanced Technology (IJEAT)*, *6*(6).
24. Aldarwish, M. M., & Ahmad, H. F. (2017, March). Predicting depression levels using social media posts. In *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)* (pp. 277-280). IEEE.
25. Nadeem, M. (2016). Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384.*

S. No	Title of the Paper	Authors	Year of Publication	Journal Name	Objective (Problem addressed)	Data Set Used	Techniques Used to Detect Depression	Performance Parameters Results
1	Multi Class Depression Detection Through Tweets using Artificial Intelligence	Muhammad Osama Nusrat et al.	2024	Journal of Artificial Intelligence Research	Detecting multiple classes of depression using AI techniques applied to tweets.	Twitter Data set	CNN, RNN, SVM	Accuracy: 96% Precision: 94.5% Recall: 93.8%
2	Explainable Depression Detection Using Large Language Models on Social Media Data	Yuxi Wang et al.	2024	Computational Linguistics and Clinical Psychology Workshop	Uses large language models (LLMs) to detect depression from social media data with a focus on explainability.	Social media posts	LLMs	–
3	Hybrid Machine Learning Models to Detect Signs of Depression	Shakir Khan et al.	2024	Journal of Machine Learning Research	Combining multiple machine learning models to improve the detection of depression signs in social media text.	Tweets, Facebook Posts	Decision Trees, SVM, Neural Networks	Accuracy: 88.5% Precision: 87.2% F1-score: 86.8%
4	Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model	Vankayala Tejaswini et al.	2024	International Journal of Data Science	Enhance the accuracy of depression detection by combining NLP techniques with a hybrid deep learning model.	Twitter, Reddit Data	NLP, Hybrid of CNN & RNN	Accuracy: 90.5% Precision: 91.3% F1-score 90.7%
5	Improving Disease Detection from Social Media Text via Self-Augmentation and Contrastive Learning	Pervaiz Iqbal Khan et al.	2024	Journal of Computational Linguistic	Depression detection models through self-augmentation and contrastive learning to handle data variability.	Twitter Data	Contrastive Learning, BERT	Accuracy: 93.2% F1-score: 92.8%
6	Deep Learning for Depression Detection Using Twitter Data	Doaa Sami Khafaga et al.	2023	Journal of Social Media Analytics	Apply deep learning techniques to Twitter data to improve depression detection.	Twitter Data	CNN, LSTM	Precision: 99.15% Recall: 98.25%
7	Depression Detection in Social Media Comments Data Using Machine Learning Algorithms	Zannatun Nayem Vasha et al.	2023	Journal of Internet Services and Applications	Detect depression in social media comments using various machine learning techniques.	Facebook, YouTube comments	SVM, RF, LR	Accuracy: 85.2% Precision: 84.7% F1-score: 84.9%
8	Survey on Design Efficient Depression Detection Using Machine Learning	Mayuri P. Saraf et al.	2023	Journal of Machine Learning Techniques	Conduct a comprehensive survey on various machine learning techniques for efficient depression detection.	Twitter, Facebook, Clinical records	SVM, NB, RF, & DL	SVM: 85.3% NB: 83.7% DT: 84.2%
9	Depression Detection Using Emotional Artificial Intelligence and Machine Learning: A Closer Review	Manju Lata Joshi et al.	2022	Journal of Artificial Intelligence and Applications	Review the integration of emotional AI and machine learning techniques for depression detection.	Posts, interview recordings, and videos	NLP, SVM, DL	Precision: 88% Recall: 85% F1-score: 86%
10	AI Therapist Using Natural Language Processing	Shephali Santosh Nikam et al.	2020	International Journal of Advanced Computer Science and Applications	System for Stress Management to provide recommendation and solution for youth, especially IT professionals.	Chatterbot Corpus	NB	-
11	Depression Detection of Tweets and A Comparative Test	Rajaraman et al.	2020	IEEE Access	Analysis of techniques like TF-IDF, NB, LSTM, Logistic Regression, Linear SVM and depression through tweets.	Sentiment 140, TWINT & word2vec	TF-IDF, NB, LSTM, LR, SVM	LSTM: 87%, NB: 74.8%, Linear SVM: 76.5%
12	Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features	AlSagri et al.	2020	Computers in Human Behavior	Depression detection using activity and content features (DDACF) classification model is used.	Tweets	SVM, NB and Decision Trees	Precision: 81.5 Recall: 79.2 F1: 80.3 AUC: 0.82
13	Predicting depression using deep learning and ensemble algorithms on raw twitter data	Shetty et al.	2020	Journal of Affective Disorders	ML classifiers are employed in the twitter dataset to identify whether a person is depressed or not.	Twitter Data set	LSTM, CNN, SVC, NB, LR, RF, GB	Precision: 83.4% Recall: 82.1% F1: 82.7%
14	Depression Detection With Sentiment Analysis Of Tweets	Hemanthkumar M et al.	2019	Expert Systems with Applications	Sentiment analysis of twitter feeds and classification as positive, neutral and negative.	-	Multinomial NB, SVM	Naive Bayes: 72.99% SVM: 72%

S. No	Title of the Paper	Authors	Year of Publication	Journal Name	Objective (Problem addressed)	Data Set Used	Techniques Used to Detect Depression	Performance Parameters Results
15	Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation.	Victor, E et al.	2019	Journal of Artificial Intelligence Research	AI based Mental Evaluation system can predict whether the patient is depressed or not through deep learning.	interview questions	Multimodal Deep Learning Model	Precision: 75% Recall: 78% F1: 76%
16	Study of Depression Analysis using Machine Learning Techniques	Devakunchari Ramalingam et al.	2019	Journal of Computational Science	Various ML techniques used to detect depression analysing tweets and various social media posts.	Various ML Techniques	Various ML techniques	Accuracy: 89% Precision: 88.5% F1: 88.2%
17	The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviours.	Fonseka et al.	2019	Journal of Affective Disorders	This study includes an overview of the literature as well as the role of AI in predicting and managing suicide risk.	-	ML methods, AI and conversational agents	Prediction Accuracy: 80-86%
18	Detecting Depression in Social Media Posts Using Machine Learning.	Biradar et al.	2018	Journal of Medical Internet Research	To create training data for the system, SentiStrength analysis has been done, and a BPNN model to classify the tweets.	BPNN and SentiStrength	BPNN model	-
19	Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique	Islam et al.	2018	Journal of Biomedical Informatics	Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, and Weighted KNN were employed as KNN Classifiers.	Facebook data	Various KNNs	-
20	Review on Mood Detection using Image Processing and Chatbot using Artificial Intelligence	Thosar et al.	2018	Image and Vision Computing	Mood detection through image processing using Haar-cascade algorithm, chatbots to detect stress through text/speech input.	An image with 6000 features	Haar-cascade algorithm and Ad-boost	Precision: 79% Recall: 77.5% F1: 78.2%
21	Depression Detection and Analysis	Oak et al.	2017	Journal of Affective Disorders	Detection of depression in an individual though RBFN. Text and speech input are considered.	53 Volunteers	RBFN (Radial basis function networks)	Speech Model: 71.4% Text Model: 64.3%
22	Depression detection using emotion artificial intelligence	Deshpande et al.	2017	IEEE Transactions on Affective Computing	NLP and sentiment analysis of tweets to detect depression.	10,000 Tweets	Multinomial Naive Bayes algorithm, SVM	Precision: 76% Recall: 74% F1: 75%
23	Image Processing Techniques To Recognize Facial Emotions	A. Mercy Rani et al.	2017	Pattern Recognition	Recognizing facial emotions based on filled mouth-regions. Viola-Jones algorithm is used.	-	Neural networks, Viola-Jones algorithm, etc.	Accuracy: 95%
24	Predicting Depression Levels Using Social Media Posts	M. M. Aldarwish et al.	2017	Journal of Medical Internet Research	A system that leverages SNS as a data source and screening tool to classify users using AI based on content on SNS.	LiveJournal, Twitter and Facebook	SVM and Naive-Bayes	Accuracy: 88.1% Precision: 87.3% Recall: 86.7%
25	Identifying depression on Twitter	Nadeem et al.	2016	IEEE Transactions on Computational Social Systems	Depression is recognised by analysing large-scale records of users' linguistic histories on social media.	2.5M Tweets	Decision Trees, SVC, NB, LR, Ridge Classifier	ROC AUC: 0.83