A MENTAL HEALTH CLASSIFICATION MODEL USING NLP AND DEEP LEARNING

Dr. Vinothini A.
Assistant Professor
Vit Chennai
Chennai, India
vinothini.a@vit.ac.in

Patel Vedant Vikrambhai
Computer Science and Engineering Vit Chennai
Chennai, India
patelvedant.vikrambhai2020@vitstudent.ac.in

Abstract

The World Health Organization (WHO) estimates that depression affects over 970 million people worldwide and can be caused by a variety of mental health conditions. Depression is therefore one of the main causes of disability in the world. To address this significant worldwide issue, early detection of mental health disorders is essential. The usefulness of deep learning networks in numerous natural language processing applications has been shown by recent studies. This work aims to utilize deep learning models to classify user-provided natural language descriptions of mental health conditions into predefined groupings. The data utilized for the experiment includes the ground truth category and a text description of mental health conditions. Text processing methods from NLP were applied to the text sample and deep learning techniques were constructed. This study also delves into the classification of natural language descriptions of mental health issues using two distinct models: T5 and Long-Term Bidirectional Memory (Bi-LSTM). Leveraging the T5 model, a transformer-based architecture, sequential information is extracted from textual data to classify mental health descriptions into predefined categories. The T5 model demonstrates promising results, achieving an accuracy of 86% of the rougelsum score before overfitting, showcasing its potential for improved performance with larger datasets. However, the Bi-LSTM model also proves effective in this classification task, achieving an accuracy of 74% before overfitting. Attempts to apply transfer learning using BERT on the same dataset yielded suboptimal results, highlighting the importance of selecting appropriate models for small-scale datasets. This research underscores the efficacy of both T5 and Bi-LSTM models in mental health classification tasks and underscores opportunities for further investigation, particularly in training more advanced models on larger datasets.

Keywords: Mental Healthcare, Psychotherapy, Deep Learning, Natural Language Processing

1. Introduction

To address this significant worldwide issue, early detection of mental health disorders is essential. Kesebeck et al. (2018) reported that a telephone survey conducted in Germany revealed that self-stigmatization and shame, as opposed to perceived stigma and adverse reactions, were the main reasons people did not seek psychiatric therapy for depression.

However, patients can now employ a range of application-based and online diagnostic tools to get a diagnosis from the comfort of their own homes thanks to recent technological breakthroughs, particularly in the field of natural language processing (NLP). The goal of NLP is, a topic at the confluence of computer science, artificial intelligence (AI), and linguistics, is to enable computers to interpret, analyze, and approximate human speech. Many chatbots that conduct mental health assessments through natural conversation are accessible as Android/iOS apps or websites. Examples of these chatbots include Woebot, Wysa, Joyable, and Talkspace. However, new technologies offer early identification and preventative treatments to prevent mental health problems from getting worse. In this study, neural networks were used to classify patients' mental health problems into discrete categories after natural language processing (NLP) techniques were applied to their explanations of the problems. This kind of classification system helps patients identify their problem areas and select the best therapist for their individual needs.

Nicolas Bertagnolli used CounselChat.com to access the online repository to gather the data for this study (Bertagnolli, 2020). The information is conversations from therapy sessions in which a patient discloses a mental health issue and receives therapeutic advice from a certified therapist. An acceptable hypothesis and the goals and objectives of the investigation will be included in the report. After that, it will look at relevant earlier studies on the problem at hand as well as the methods being used now. The report will then explain the research strategy used to resolve the problem. After that, the study results will be discussed and their relevance will be provided in the paper.

2. Literature Survey

[1] RoBERTa-Based Categorization of Mental Health Conditions on Social Media. In this study, social media posts that highlight mental illness especially those from Reddit are categorized using the RoBERTa model. The study offers a brand-new dataset made up of titles and posts from several Reddit subreddits that discuss mental health conditions like PTSD, ADHD, bipolar disorder, despair, and anxiety. When trained on posts and titles, the RoBERTa model demonstrated good accuracy, with an F1 score of 0.89. This method's benefits include the efficient application of a Transformer-based architecture for examining mental health and emotions in writings shared on social media. Additionally, the study shows how social media data might support professional practices in mental health prognosis. The study does, however, have certain shortcomings.

[2] Use of machine learning algorithms for psychiatric problem detection and diagnosis. This study looks into the detection of psychological problems through text-based documents, reports, electronic health records, and machine learning. For precise diagnosis, the suggested solution combines patient-specific input with machine learning algorithms. Important findings include the successful diagnosis of psychological diseases using Regularized General Linear Models (GLMs), Support Vector Machines, and Single-layer Artificial Neural Networks. Nonetheless, it encounters obstacles including the intricacy of mental health conditions, fluctuations in indications, and the requirement for meticulous evaluation of data confidentiality and protection.

[3] Cognitive distortions in mental health texts are automatically detected and classified. It focuses on classify and identify cognitive distortions found in mental health literature using machine learning. mixing crowdsourced datasets with actual online therapy programs. Support vector machines (SVM), logistic regression, random forests, gradient-boosted trees, recurrent neural networks (RNN), convolutional neural networks (CNN), and BERT are a few of the methods that are employed. Observations and results show that logistic regression outperformed other models due to the specific nature of cognitive distortions and their expression in language. The study reports high accuracy in detecting cognitive distortions but faces challenges in classifying them into specific categories due to the complexity and co-occurrence of these distortions in real-world settings. The advantages of this approach include its potential to assist in online mental health services by providing instant feedback on distorted thinking. However, the challenge lies in accurately typing specific cognitive distortions and managing the variability in real-world mental health texts.

[4] A Novel Co-training base Classification of Mental Illnesses Using Social Media. It examines the classification of Using information from Reddit posts, researchers examined mental diseases like ADHD and bipolar disorder, sadness, and anxiety. The study employs a Co-training approach, which is a type of semi-supervised learning, and utilizes classifiers like Naïve Bayes, Random Forest, and Support Vector Machine. The research demonstrates that the Co-training approach improves classification accuracy compared to standard classifiers. It effectively uses posts and comments from Reddit for mental illness classification, with the Co-training approach demonstrating improved results for a range of mental health issues in terms of recall, accuracy, and F-measure. A key advantage of this approach is its ability to leverage unlabeled data for improved model performance. However, the study acknowledges limitations in generalization as it relies on data from a single platform (Reddit) and uses default parameters for classifiers, which may affect the overall effectiveness The paper reports that the Co-training approach achieved an accuracy of 91.3% for Anxiety, 90.1% for Depression, 89.5% for Bipolar Disorder, and 88.7% for ADHD. These results indicate a significant improvement in classification accuracy for mental health conditions using the Co-training method compared to traditional classifiers.

[5] Artificial Mental Health Record Generation and Assessment for Natural Language Processing. This paper describes a process for creating synthetic clinical papers, focusing on mental health records, and evaluates their utility for natural tasks using language processing (NLP). Findings: The study discovered that training with simulated data produced classification outcomes that were on par with those obtained using original data. In particular, the model known as Convolutional Neural Network (CNN) was the best-performing, showing significant improvements over other models. The F1-scores for the CNN model using genuine data averaged 0.48, while the top+meta artificial data method achieved an average F1-score of 0.43. However, a limitation is the challenge of ensuring that the artificial data does not retain sensitive information from the original data.

[6] Predicting Social Anxiety, It focuses on utilizing machine learning algorithms to forecast social anxiety treatment results. disorder based on email conversations between patients and therapists. Results: The study achieved promising results, with an Area Under the Curve (AUC) of 0.83 when predicting therapy outcomes halfway through the treatment using email data and socio-demographic information. When using the full data set (entire treatment period), the precision achieved was 0.78. Advantages include the innovative use of real-world, text-based patient-therapist interactions for predicting treatment outcomes, offering a potential tool for early intervention in social anxiety treatments. However, the study acknowledges limitations due to the small sample size and the specificity of the dataset to social anxiety disorder, which might limit the generalization of the findings.

[7] Using machine learning techniques to analyze financial text sentimentally in order to monitor the mental health of the public. It looks at how machine learning is used to analyze public sentiment based on financial texts for monitoring mental health. The techniques used include AdaBoost, Single Layer Convolutional Neural Network (SLCNN), and Support Vector Machine (SVM). Results:- The study demonstrates that the Single Layer Convolutional Neural Network (SLCNN) outperforms other techniques, achieving a classification accuracy of 93.9%. The Support Vector Machine (SVM) and AdaBoost techniques also showed significant performance with classification accuracies of 67.7% and 76.1% respectively. The advantages of study include the use of machine learning for analyzing public sentiment in financial texts, contributing to understanding public mental health. A limitation is the study's focus on financial texts, which may not fully represent the broad spectrum of factors influencing the public.

[8] This paper proposes a framework that includes Using social media posts, bi-directional long short-term memory (Bi-LSTM) and BERT are used to analyze and identify symptoms of anxiety and depression. Results:- The system achieved an accuracy of 98% using a knowledge distillation technique. This high accuracy demonstrates the effectiveness of combining advanced NLP techniques for mental health prediction. The approach effectively analyzes social media data to identify mental health issues, demonstrating the potential of deep learning in public health monitoring. However, the paper's focus on specific social media platforms may limit its applicability to broader datasets or different contexts.

[9] Automatic Determination of Depression Level using Visual Input. This study uses pictures and videos as input and convolutional neural networks (CNNs) to detect depression levels using the Beck Depression Inventory-II (BDI-II). With CNN, the machine learning model's accuracy was 66.45%. The methodology involves preprocessing and feature extraction from video frames, followed by predicting the degree of depression based on the relationship between the emotion vector and the BDI-II assessment score. Advantages include the innovative use of visual data for depression detection, which can be valuable for remote monitoring and early intervention. However, the study faces limitations in terms of lower accuracy and potential biases due to facial expressions' variability.

[10] The study, examines how deep transfer learning, namely BERT, is used to predict emotional valence in mental health literature. The study compares different models, including logistic regression and RNNs, with BERT showing superior performance. Results indicate that the BERT model, trained on social media data and fine-tuned on mental health texts, outperformed other models in terms of F1 scores, recall, and precision. In comparison to other models, the BERT model achieved higher weighted F1 score (85%), macro F1 score (76%), and accuracy (85%). This approach's advantage lies in its ability to accurately analyze and classify emotional valence in mental health texts. Nonetheless, the caliber and representativeness of the training data determine how well the model performs.

[11] Study on Mental Disorder Detection using Social Media Mining looks at data from social media to find mental health problems. The study implements computational linguistic processes with features like SenticNet's emotional state dimensions, self-reference, and mental disorder word count. Results: The method identified 8,105 tweets, categorizing them into three levels of depression: high (1,028 tweets), moderate (1,073), and low (1,605). Advantages include the novel application of text mining for mental disorder detection on social media, offering a new approach to mental health monitoring. However, the study's reliance on self-declared diagnoses and social media data may affect its accuracy and generalizability.

[12] The use of predictive modeling to examine depression and mental health problems using data from social media is examined in the article "Predictive modeling techniques: An analysis of depression and impaired mental health from

social media platforms." The research focuses on Instagram data, using machine learning models to analyze text and visual features from user posts. The study implemented a Multimodal approach, which achieved around 60% accuracy in predicting users' mental health states. This approach involved a combination of text and visual feature analysis, with a focus on semantic responses and image characteristics. Advantages include the novel use of a multimodal approach to analyze both text and visual social media data for mental health analysis. However, the study faced limitations due to the discontinuation of Instagram's API, which made data collection challenging, and the accuracy was moderate.

[13] A unique chatbot system for mental counseling is shown in A counseling service that uses emotional response generation called The ChatBot Feels You. This system is designed to understand and generate responses based on user emotions using advanced natural language processing methods. Results:- The paper does not provide specific numerical results such as accuracy percentages. Instead, it focuses on the conceptual framework and the implementation of the chatbot system, emphasizing its ability to recognize and generate emotional responses for counseling purposes. Advantages include the potential for continuous and real-time emotional monitoring in a counseling context, enhancing the effectiveness of mental health support. However, the paper does not address the specific challenges of implementing such a system in real-world scenarios or its effectiveness in diverse patient populations.

[14] Sentiment Analysis using LSTM focuses on categorizing movie reviews with Long Short-Term Memory (LSTM) networks from the IMDB dataset. The LSTM model achieved an accuracy of 86.85% in sentiment classification, indicating its effectiveness in handling sequential text data like movie reviews. The paper highlights LSTM's capability to process long sequences of text data, offering improvements over traditional machine learning methods in sentiment analysis tasks. But it's crucial to remember that the paper's scope is limited to movie reviews, which may not fully represent broader sentiment analysis applications.

[15] The included After a comprehensive screening process, the key findings from each study were condensed into unique tables that included details on the authors, the year the study was published, particularly mental health issues, demographics, and data types that were collected. The paper also discusses the results of the study selection process, topics and population characteristics, and the citations and citations of the included studies. The review's advantages and disadvantages are discussed in the discussion section, which also emphasizes the lack of quantitative comparisons between the chosen studies and the methods.

3. Creative Mechanism Design Methodology

The methodology involves implementing NLP and Deep Learning algorithms, utilizing natural language processing for accurate detection and classification. Initial Steps include dataset collection and preprocessing ensuring diversity in mental health issues and classification Algorithm training follows employing deep learning frameworks and incorporating the text data.

3.1 Proposed System

•**Data collection and annotation:-** The next step is to collect a large dataset of text data that will be used to train the T5 model. This involves collecting data from a variety of sources, such as social media, online forums, and healthcare records, and annotating it with labels that indicate the presence or absence of each mental health issue.

•**Preprocessing:-** The collected text data must be preprocessed to remove noise, standardize language use, and convert it into a format that can be used by the T5 model. This may involve techniques such as tokenization, stemming, and lemmatization.

•**Model development**:**-** The T5 model is then developed using the preprocessed text data. This involves selecting appropriate hyperparameters, training the model on the labeled data, and evaluating its performance using appropriate metrics such as precision, recall, and F1 score.

•**Deployment:-** Once the T5 model has been developed and evaluated, it can be deployed in a production environment. This may involve integrating the model into an existing mental health screening tool, developing a standalone application, or integrating it into an electronic health record system.

3.2 Proposed System Diagram

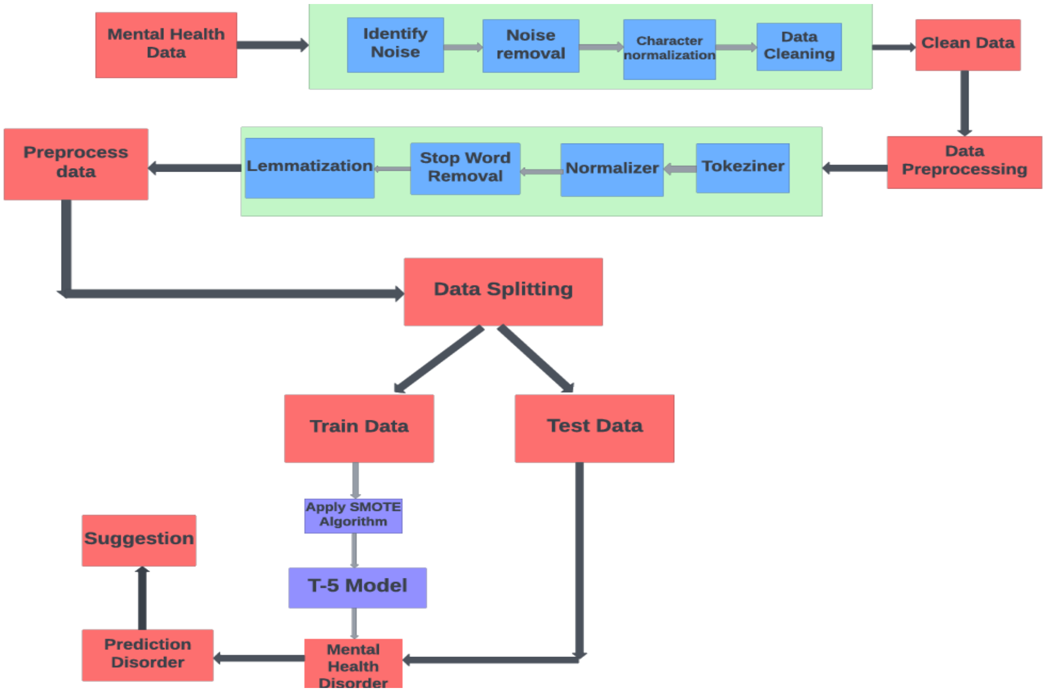

Figure 3.1 Proposed System Diagram

3.3 MODULES

1. Data Collection and Preprocessing

- Word embedding
- Tokenization
- Stemming
- Lemmatization
- Part of speech tagging
- Normalizer
- Stop words Removal

2. Hyperparameter Tuning

3. Performance Evaluation

1. Data Collection and Preprocessing

•Data Collection:- Gather a diverse dataset containing text samples related to mental health. This can include forums, social media posts, clinical notes, etc.

•Text Cleaning:- Preprocess the text data by removing irrelevant information, and special characters, and performing tasks like stemming or lemmatization.

•Tokenization:- Break down the text into individual words or tokens.

Word Embeddings:- Word Embedding Layer: Convert words into numerical vectors to represent semantic relationships. Techniques like Word2Vec, GloVe, or pre-trained embeddings (e.g., Word2Vec, GloVe, or embeddings from transformer models like BERT) can be used.

Tokenization:- Tokenization breaks text into smaller parts for easier machine analysis, helping machines understand human language. Tokenization, in the realm of NLP and machine learning, refers to the process of converting a sequence of text into smaller parts, known as tokens.

Stemming:- Stemming is a technique used to extract the base form of words by removing affixes from them. It is just like cutting down the branches of a tree to its stems. For example, the stem of the words eating, eats, and eaten is eat.

Lemmatization:- Lemmatization is a text pre-processing technique used in natural language processing (NLP) models to break a word down to its root meaning to identify similarities. For example, a lemmatization algorithm would reduce the word better to its root word, or lemme, good.

Part of speech tagging:- The process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context.

Normalizer:-Normalization helps you keep your database organized. It reduces the duplication of data and improves its integrity. In its first normal form, the relation consists of unique values with no-repeat values. The second normal form eliminates all partial dependencies.

Stop words Removal:- The idea is simply removing the words that occur commonly across all the documents in the corpus. Typically, articles and pronouns are generally classified as stop words.

2. Hyperparameter Tuning

Grid Search or Random Search: Experiment with different hyperparameter values (e.g., learning rate, dropout rate, LSTM units) to optimize model performance.

3. Evaluation Metrics

Accuracy, Precision, Recall, and F1 Score:- Evaluate the model's performance using appropriate metrics for classification tasks.

Confusion Matrix:- Analyze false positives, false negatives, true positives, and true negatives.

4. Result and Discussion

(1) LSTM training

LSTM Model Result Table

Table 4.1 Training an LSTM model using different epoch and batch size

Epoch	Batch Size	Training Accuracy	Training Loss	Precision	f1_score
100	128	100%	2.2127	0.959514	0.959237
10	100	100%	1.3726	0.914352	0.947977
20	50	98.96%	2.2703	0.962406	0.960393
50	100	99.86%	1.1138	0.868132	0.958080
130	128	100%	4.1261	0.904569	0.657971

The results of training an LSTM model with various batch sizes and epoch configurations are shown in the table 4.1. With the "Epoch" column representing the number of iterations over the dataset and the "Batch Size" column indicating the amount of data points processed in each training phase, each row represents a distinct training scenario. The proportion of correctly identified cases during training is shown in the "Training Accuracy" column, while the model's inaccuracy in predicting outputs in comparison to actual values is measured in the "Training Loss" column. Higher values in the precision and F1 Score measures indicate better performance and accuracy of the model's predictions overall.

Table 4.2 Testing an LSTM model using different epoch and batch size

Epoch	Batch Size	Val. Accuracy	Val. Loss	Val. Precision	Val. f1_score
100	128	74.63%	3.3631	0.710145	0.666667
10	100	73.18%	3.7837	0.746377	0.724638
20	50	67.15%	3.6300	0.654589	0.702899
50	100	70.77%	2.5812	0.714976	0.705314
130	128	76.11%	2.7621	0.850123	0.661836

Fifty epochs were used to train the model. Training accuracy rose steadily until it attained 100%. Nonetheless, the test's accuracy leveled off at roughly 70%. Considering the magnitude of the dataset, this is an excellent outcome. Since the network will have more word tokens to rely on, the test accuracy may rise even further with additional data. Table 4.2 displays the graph for the loss plot. This picture illustrates how, shortly before the 20th epoch, the model begins to overfit to the training set.

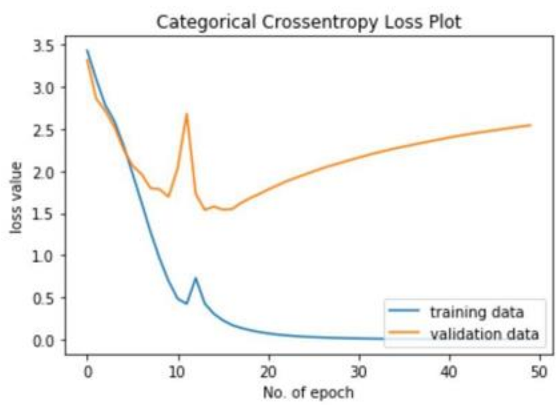

Figure 4.1 Training an LSTM model using a categorical cross-entropy loss plot

The graphs are displayed in Figure 4.1 after the findings were subjected to several categorizations.

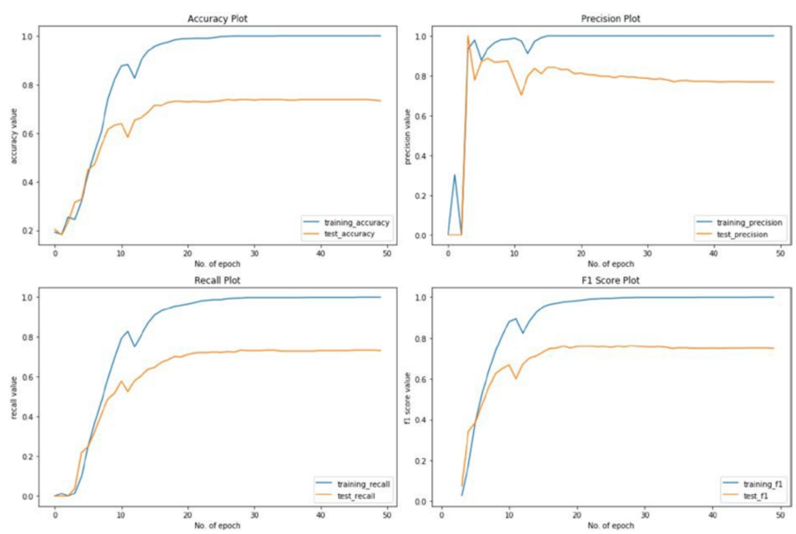

Figure 4.2 LSTM model training classification metrics

The model assumes a smooth convergence around the 20th epoch, according to the plots displayed in Figure 4.2. Plots of precision, recall, and F1 score provide additional evidence that the model is capable of accurately classifying data.

(2) T5 training

T-5 Model Result Table

Table 4.3 Training and Testing T5 model using different epoch

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	Rougel	RougeSum
10	No log	0.614593	82.21%	53.33%	81.73%	81.51%
30	0.005400	0.944204	82.91%	53.19%	81.94%	81.86%
50	0.001600	0.987690	83.01%	53.79%	82.57%	82.53%
100	0.040000	0.995669	86.11%	57.33%	84.79%	85.77%
150	0.000231	1.000000	79.90%	61.99%	82.31%	86.31%

The performance of a T5 model trained over different epochs is shown in the table 4.3. It displays the validation loss (error during validation), training loss (difference between the model's predictions and the real data during training), and Rouge scores (metrics assessing the quality of text summarization produced by the model) for each epoch. The longest common subsequence is measured by RougeL, the overlap of bigrams is measured by Rouge2 and the overlap of unigrams by Rouge1, and the average of these three measures is called RougeSum. The results show that the Rouge scores show constant performance in producing summaries that closely resemble reference summaries, and this model got 86.11%.

When the model ran with 100 epochs it got 0.04 training loss, 0.99 validation loss, 86.11% rouge1 score, 57.33% rouge2 score, and 85.77% rougesum score which is better the accuracy of my project.

(3) BERT training

BERT Model Result Table

Table 4.4 Training BERT model using different epoch and batch size

Epoch	Batch Size	Training Accuracy	Training Loss	Precision	f1_score
130	100	83.90%	0.150049	0.896552	0.998440
50	50	94.80%	0.079595	0.833333	0.990596
20	30	98.89%	0.012880	1.000000	0.954397
10	128	76.73%	0.050514	0.995749	0.973348
100	128	99.79%	1.251614	0.979971	0.824308

To investigate the effects of different combinations of epochs and batch sizes on model performance, this table 4.4 presenting the results of BERT model training was examined. For example, the training accuracy was 83.90% at 130 epochs with a batch size of 100, and the training loss was 0.150049. The measures for precision and F1 score were likewise excellent, coming in at 0.896552 and 0.998440, respectively. On the other hand, training accuracy increased to 99.79% at 100 epochs with a batch size of 128; nevertheless, the model had a greater training loss (1.251614), lesser precision (0.979971), and an F1 score (0.824308). These variants show how sensitive the BERT model is to various training setups, with some configurations producing better performance metrics and accuracy than others.

Table 4.5 Testing BERT model using different epoch and batch size

Epoch	Batch Size	Val. Accuracy	Val. Loss	Val. Precision	Val. f1_score
130	100	68.03%	2.422356	0.762148	0.740000
50	50	77.73%	2.186854	0.818627	0.741294
20	30	71.79%	1.922006	0.835165	0.727960
10	128	71.98%	3.323464	0.786096	0.732997
100	128	54.21%	2.496670	1.000000	0.747141

Regarding the second table 4.5, which assesses the effectiveness of the BERT model on test data, comparable fluctuations were noted for various combinations of epoch and batch size. For example, the model scored 68.03% validation accuracy with 2.422356 validation loss at 130 epochs and 100 batch size. The F1 score was 0.740000, and the precision was 0.762148. As a result of overfitting, the validation accuracy decreased to 54.21% with 100 epochs and a batch size of 128 while the validation precision increased to 1.000000. These findings highlight how crucial it is to choose the right batch size and epoch combinations in order to guarantee the best.

Figure 4.3 Training BERT model using a categorical cross-entropy loss plot

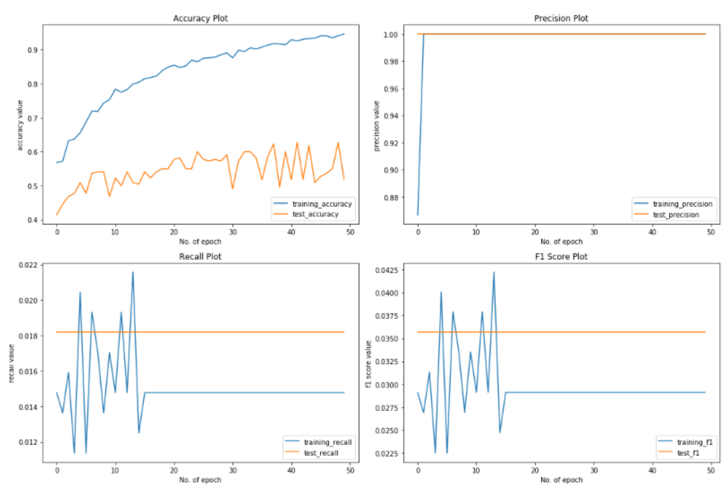

Figure 4.4 BERT model training classification metrics

The graphs are displayed in Figure 4.4 after the findings were subjected to some categorizations.

(4) Bi-LSTM training

BI-LSTM Model Result Table

Table 4.6 Training BI-LSTM model using different epoch and batch size

Epoch	Batch Size	Training Accuracy	Training Loss	Precision	f1_score
100	128	100%	0.002321	1.000000	0.999481
10	100	99.89%	0.003173	0.998962	0.822994
20	50	99.79%	0.007669	0.936656	0.990566
50	100	100%	0.017871	0.992731	0.998960
130	128	90.13%	1.628857	1.000000	1.000000

The Bi-LSTM model's training results are shown in the first table 4.6 for a variety of batch sizes and epoch combinations. "Epoch" indicates the number of training iterations, while "Batch Size" indicates the number of data samples handled in each training step. Each row represents a distinct training scenario. For example, the model achieved 100% training accuracy with a low training loss of 0.002241 at 100 epochs and a batch size of 128. The parameters for precision and F1 score were likewise excellent, coming in at 1.000000 and 0.999481, respectively. On the other hand, in 128 batches and 130 epochs, the accuracy remained high at 90.13%, but the training loss climbed dramatically to 1.628857, suggesting that overfitting may have occurred.

Table 4.7 Testing BI-LSTM model using different epoch and batch size

Epoch	Batch Size	Val. Accuracy	Val. Loss	Val. Precision	Val. f1_score
100	128	73.36%	2.541599	0.768448	0.749380
10	100	66.43%	1.994534	0.782383	0.751561
20	50	58.35%	1.727211	0.873469	0.760456
50	100	72.88%	2.031993	0.770408	0.626357
130	128	54.96%	3.315263	0.793103	0.702247

Regarding the second table 4.7 which assesses the performance of the Bi-LSTM model on test data, a comparable pattern of fluctuations was noted for various batch size and epoch configurations. For instance, the model achieved a validation accuracy of 73.36% with a validation loss of 2.541599. This was done at 100 epochs and a batch size of 128. The F1 score was 0.749380, and the precision was 0.768448. On the other hand, the validation accuracy decreased to 54.96% with 130 epochs and a batch size of 128. This may indicate problems with generalization. These findings highlight how crucial it is to choose the right batch size and epoch combinations in order to maximize the Bi-LSTM model's performance on unknown data.

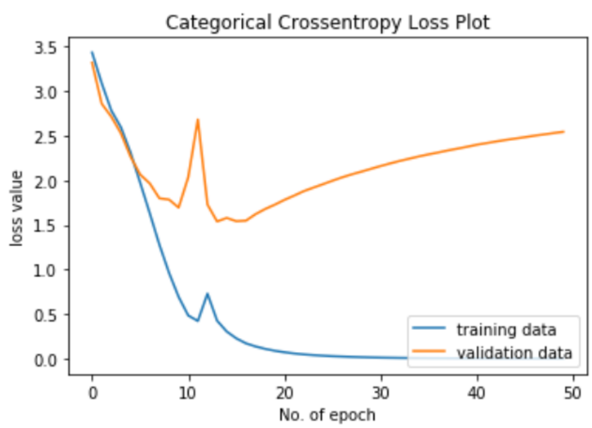

Figure 4.5 Training BI-LSTM model using a categorical cross-entropy loss plot

Above figure 4.5 describe how Bi-LSTM works with loss value against No of epochs.

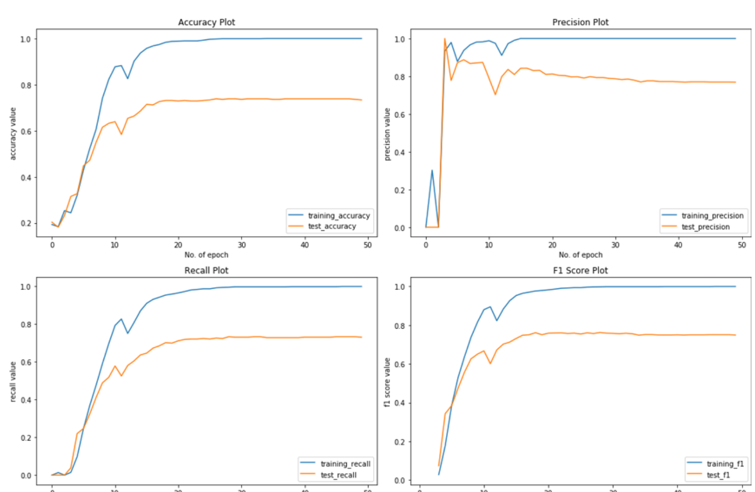

Figure 4.6 BI-LSTM model training classification metrics

5. Conclusions

Our project is to develop an algorithm that can be used by natural language data analysis systems to understand mental health issues in patients. All users would need to do is describe their problem to the system as they would to a therapist, and it would use NLP techniques to process that explanation and categorize the user. A system like this has several applications. A counseling therapy provider may choose to have their patient complete an online initial evaluation to help them automatically match the patient with an appropriate therapist. This suggests that providers might be able to lower the cost of first assessments. Additionally, it might be applied to mental health counseling software, which analyzes patients' natural language descriptions of their problems to try and unite them. Categorizing every discussion that patients have on the application would also make it easier for the application to keep track of all the problems that patients have.

6. Future Work

There is always room for improvement in the performance of the T5 model used for mental health issue classification. This may involve exploring new architectures, such as hybrid models that combine T5 with other deep learning techniques, or improving the feature extraction process. Additionally, Future work should focus on creating models that are more resilient and culturally and linguistically sensitive. Another important area of future work is cross-disciplinary research that brings together experts in NLP, mental health, and related fields. Such research can explore novel applications of T5-based approaches to mental health issues, such as the early detection of mental health issues, predicting suicide risk, and developing personalized treatments based on patient data. This type of research can help bridge the gap between technology and clinical practice and provide innovative solutions to mental health issues. Overall, the future work for The use of NLP to classify mental health concerns using T5-based approaches is a promising method that has numerous prospects for future research and advancement. As NLP technology continues to advance, T5-based approaches will likely become more widely used in mental health research and clinical practice.

Exploring Model Architectures

• Investigating novel architectures that combine BERT, Bi-LSTM, and LSTM models with other deep learning techniques to improve performance and address specific challenges in mental health classification tasks.

• Experimenting with transformer-based architectures beyond BERT, such as Roberta, ALBERT, and GPT models, to leverage advancements in natural language understanding and generation.

Enhancing Multimodal Capabilities

• Integrating multimodal data sources, such as text, images, and audio, to develop more comprehensive models for mental health classification.

• Exploring techniques for multimodal fusion and attention mechanisms to effectively combine information from different modalities and improve classification accuracy.

Adapting to Multilingual and Multicultural Contexts

• Extending models to support multiple languages and cultural contexts to ensure their applicability and effectiveness in diverse populations.

• Investigating techniques for cross-lingual transfer learning and domain adaptation to leverage data from different languages and cultures without the need for extensive labeled data.

Addressing Ethical and Privacy Considerations

• Developing methods for privacy-preserving training and inference to protect sensitive patient information while still enabling effective mental health classification.

• Conducting research on bias mitigation and fairness in model predictions to ensure equitable outcomes across different demographic groups.

Advancing Clinical Applications

• Developing techniques for model explainability and interpretability to provide insights into the decision-making.

References

[1] B. Shickel, M. Heesacker, S. Benton and P. Rashidi, "Automated Emotional Valence Prediction in Mental Health Text via Deep Transfer Learning," 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 2020. (IEEE)

[2] Dongkeon Lee, Kyo-Joong Oh, and Ho-Jin Choi, "The chatbot feels you - a counseling service using emotional response generation,(BigComp), Jeju, Korea (South), 2021. (IEEE)

[3] Syarif, N. Ningtias and T. Badriyah, "Study on Mental Disorder Detection via Social Media Mining," 2019 4th International Conference on Computing, Communications and Security (ICCCS).2022(IEEE)

[4] Y. Aryal, A. Maag and N. Gunasekera, "Application of Machine learning algorithms in diagnosis and detection of psychological disorders.(IEEE)-2020.

[5] Tanya Nijhawan, Girija Attigeri & T. Ananthakrishna,Stress detection using natural language processing and machine learning over social interactions, 2022(Springer).

[6] Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor,1 Philippe Lenca, Romain Billot, Taylor C Ryan, Jonathan Marsh, Jordan DeVylder, Michel Walter,1, Sofian Berrouiguet, Christophe Lemey,Machine Learning and Natural Language Processing in Mental Health,2021(Jounral of medical internet research).

[7] Young, T., Hazarika, D., Poria, S. and Cambria, E., 2018. Recent Trends in Deep Learning Based Natural Language Processing, ArXiv, 2020(IEEE).

[8] M. Hoogendoorn, T. Berger, A. Schulz, T. Stolz and P. Szolovits, "Predicting Social Anxiety Treatment Outcome Based on Therapeutic Email Conversations," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 5, pp. 1449-1459, Sept.(2022)

[9] Tholusuri, A., Anumala, M., Malapolu, B. and Lakshmi, G.J., 2019. Sentiment Analysis using LSTM, International Journal of Engineering and Advanced Technology, 8(6S3), pp. 1338 1340.(IJEAT)

[10] Su, M., Wu, C., Huang, K and Hong, Q., 2018. LSTM-based Text Emotion Recognition using Semantic and Emotional Word Vectors, First Asian Conference on Affective Computing and Intelligence Interaction,2020(ACII Asia).

[11] Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space, ArXiv,2020 (IEEE).

[12] K. M. Hasib et al., "MCNN-LSTM: Combining CNN and LSTM to Classify Multi-Class Text in Imbalanced News Data," in IEEE Access, vol. 11, pp. 93048-93063, 2023, doi: 10.1109/ACCESS.2023.3309697. (IEEE).

[13] B. C. Loftness et al., "Toward Digital Phenotypes of Early Childhood Mental Health via Unsupervised and Supervised Machine Learning," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-4.

[14] Bigi, B., 2014. A multilingual text normalization approach. Human Language and Technology Challenges for Computer Science and Linguistics, vol. LNAI-8387, pp. 515-526(IEEE) -2020.

[15] H. K. Marrapu, B. Maram and P. Reddi, "New Analytic Framework of Public Mental Health Prediction Using Data Science," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2022, pp. 1-6, (IJEAT).

[16] Neha S, Shekar P H C., Kumar K S., and Vg A, ―Emotion recognition and depression detection using deep learning,‖ pp. 3031–3036, 2020.(Springer)

[17] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 23(5), 649–685.(IEEE)

[18] Abha Tewari, Amit Chhabria ,Ajay Singh Khalsa, Sanket Chaudhary ,Harshita Kanal,‘A Survey of Mental Health Chatbots using NLP,2021.(IEEE)

[19] Tanya Nijhawan, Girija Attigeri & T. Ananthakrishna,Stress detection using natural language processing and machine learning over social interactions, 2022.(IEEE)

[20] Young, T., Hazarika, D., Poria, S. and Cambria, E., 2018. Recent Trends in Deep Learning Based Natural Language Processing, ArXiv, 2020.

[21] Webster, J.J. and Kit, C., 1992. Tokenization as the initial phase in NLP, Proceedings of the 14th conference on Computational Linguistics, August, pp, 1106-1110.

[22] Tholusuri, A., Anumala, M., Malapolu, B. and Lakshmi, G.J., 2019. Sentiment Analysis using LSTM, International Journal of Engineering and Advanced Technology, 8(6S3), pp. 1338- 1340.

[23] Dataset https://counselchat.com/