INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

Natural Language Processing (NLP)-Based Detection of Depressive Comments and Tweets: A Text Classification Approach

Jose C. Agoylo Jr., Kim N. Subang, Jorton A. Tagud

BSIT, Southern Leyte State University – Tomas Oppus Campus, Southern Leyte, Philippines

DOI : https://doi.org/10.51583/IJLTEMAS.2024.130606

Received: 14 June 2024; Revised: 28 June 2024; Accepted: 02 July 2024; Published: 12 July 2024

Abstract: Depression is a major mental health problem that affects millions globally, causing significant emotional distress and impacting quality of life. With the pervasive use of social media platforms, individuals often express their thoughts and emotions through online posts, comments, and tweets, presenting an opportunity to study and detect depressive language patterns. This research utilized the dataset from Kaggle between December 2019 and December 2020, which originated largely from India. This paper presents a novel approach for detecting depressive sentiment in online discourse using Natural Language Processing (NLP) and machine learning techniques. The study aims to develop an automated system capable of accurately identifying depressive comments and tweets, facilitating early intervention and support for individuals potentially struggling with mental health challenges. The proposed methodology will be rigorously evaluated using standard performance metrics, including precision, recall, F1- score, and ROC curve. The study will also conduct qualitative analyses to gain insights into the types of textual patterns and linguistic cues most indicative of depressive sentiment. The results of our study are promising, with a maximum validation accuracy of 0.88 demonstrating the model's ability to classify depressive and non-depressive comments and tweets accurately. The outcomes of this research have significant implications for mental health monitoring and intervention strategies. By accurately detecting depressive sentiment in online discourse, healthcare professionals and support services can proactively reach out to individuals exhibiting potential signs of depression, fostering early intervention and improving overall mental health outcomes.

Keywords: Depression, F1-score, long short-term memory (LSTM), mental health, natural language processing (NLP).

I. Introduction

Depression is a significant mental health disease marked by chronic sadness, loss of interest, and a range of emotional and physical problems. According to the World Health Organization, depression affects over 300 million people globally and is a leading cause of disability worldwide. Early identification and intervention are essential for managing depression and preventing further deterioration of mental health. As social media platforms have grown in popularity, individuals increasingly use these channels to express their thoughts, emotions, and experiences, often reflecting their mental state. They introduced BERT (Bidirectional Encoder Representations from Transformers), a groundbreaking model that improved the state-of-the-art in various NLP tasks through bidirectional training of transformers [7]. However, [19] addressed the limitations of BERT by leveraging the best of both autoregressive and autoencoding approaches, setting new records in NLP benchmarks. Thus, [15, 5], proposed the Text-To-Text

Transfer Transformer (T5) model, demonstrating the versatility of framing all NLP tasks as text- to-text problems. According to [3, 17], the proposed GPT-3, with 175 billion parameters, shows impressive few-shot learning capabilities across various NLP tasks without needing task-specific fine-tuning but [4, 9, 11] developed a more sample-efficient pre-training method by training a discriminator to distinguish real tokens from corrupted ones. This online content provides a rich source of data for studying and understanding depressive language patterns. Natural Language Processing (NLP) techniques, combined with machine learning algorithms, offer a powerful approach to analyzing and classifying this textual data, enabling the detection of depressive language and potentially identifying individuals at risk of depression. As LSTMs have been widely adopted, researchers have also focused on improving their interpretability and explain ability [1, 13].

The proposed research seeks to produce an effective NLP text classification model capable of accurately distinguishing depressive comments or tweets from non-depressive ones. By leveraging a large dataset of social media content, the researchers will explore the use of pre-trained word embeddings and a Long Short-Term Memory (LSTM) based deep learning architecture to optimize the classification performance. The ultimate goal is to design a tool to help mental health experts. and support services in identifying individuals potentially struggling with depression, enabling timely intervention and support.

Research Objectives

The researchers aimed to achieve the following objectives;

1. To develop an NLP Text Classification Model;
2. To leverage social media datasets; and;
3. To investigate the performance of LSTM-based Deep Learning Architectures.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

Conceptual Framework of the Study

The conceptual framework presented in the image outlines a comprehensive approach for analyzing large datasets of online comments and tweets from multiple social media platforms, to detect depressive sentiments and enable mental health research and performance analysis. this conceptual framework combines the power of large datasets, advanced NLP techniques, and state-of-the-art machine learning models to tackle the complex challenge of detecting depressive sentiments in online expressions. By leveraging this framework, researchers can gain valuable insights into mental health and human behavior, potentially leading to better understanding, support, and interventions for people suffering from depression or other mental health.

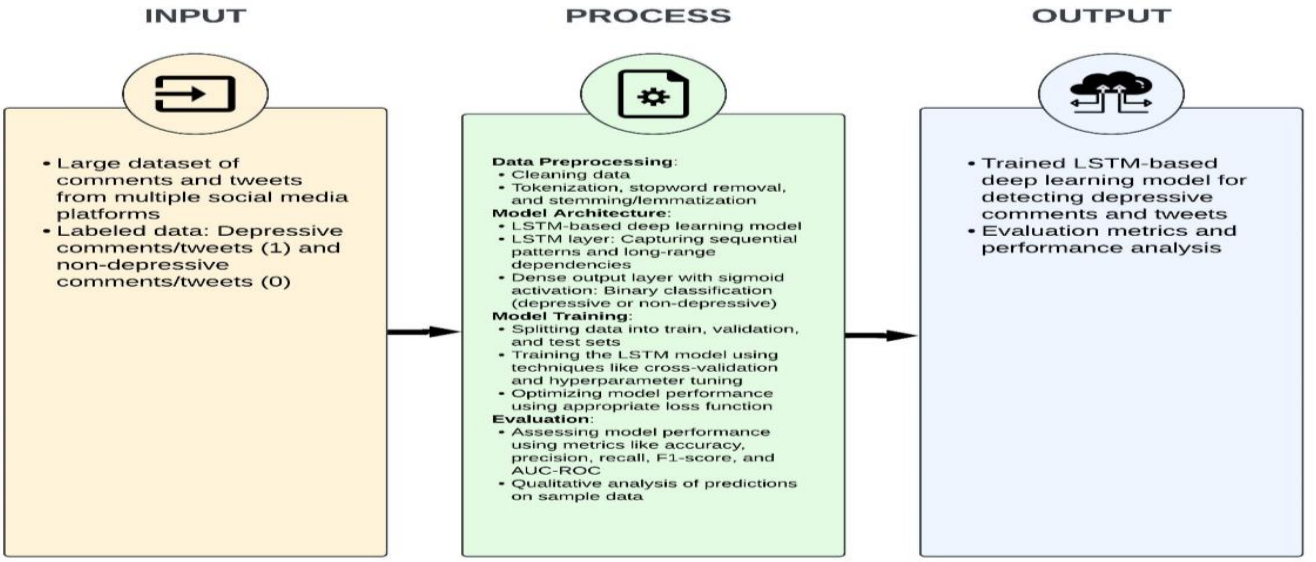

Fig.1 Conceptual Framework

II. Methodology

Data Collection and Preprocessing

The research utilized a dataset from Kaggle, focusing on depressive and non-depressive tweets posted between December 2019 and December 2020, primarily from India and neighboring regions. Tweets were selected based on the top 250 most commonly used negative and positive words, identified through SentiWord and various academic publications. To develop a robust and generalizable model, the researchers collect a large dataset of comments and tweets from multiple social media platforms, including Twitter, Reddit, and mental health forums. The dataset will consist of both depressive and non-depressive content, ensuring a balanced representation of the two classes.

Ethical considerations will be paramount during data collection, ensuring compliance with platform policies and privacy regulations. Personally identifiable information will be removed, and appropriate measures will be taken to protect user anonymity.

The collected data will undergo a comprehensive preprocessing pipeline, including cleaning data, irrelevant tokens, correcting spelling and grammatical errors, tokenization, stop word removal, and stemming or lemmatization. This preprocessing step aims to clean and standardize the textual data, improving the quality and consistency of the input for the feature extraction and classification stages.

Model Architecture and Training

For the text classification task, we will employ a deep learning model architecture based on a Recurrent Neural Network (RNN) variant known as Long Short-Term Memory (LSTM). This architecture is well-suited for capturing sequential patterns and long-range dependencies in textual data, making it a promising choice for depressive language detection.

Model Components

- **Embedding Layer:** This layer will convert the input text sequences into dense vector representations using a pre-trained word embedding model. The num_unique_words parameter represents the size of the vocabulary, and the embedding dimension is set to 32.
- **LSTM Layer:** The LSTM layer will process the embedded sequences and capture the sequential patterns and dependencies within the text. The layer has 64 units, and a dropout rate of 0.1 is applied to regularize the model

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

and prevent overfitting.

- **Dense Output Layer:** The final layer is a fully connected dense layer with a single output neuron and a sigmoid activation function. This layer will produce the binary classification output, indicating whether the input text is depressive or non-depressive.

The LSTM layer plays a vital role in analyzing depressive language patterns by effectively capturing long-term dependencies in text. This capability is crucial for understanding the context and subtle nuances that may indicate depressive sentiments. The LSTM's power lies in its unique architecture, which includes a cell state and a system of gates - input, forget, and output. These components work together to selectively retain or discard information over extended sequences of text. From a mathematical perspective, the LSTM's operation revolves around the cell state Ct, which serves as the network's memory. This cell state is carefully managed through the coordinated actions of the gates. Each gate is mathematically formulated to regulate the flow of information, allowing the network to learn which information is relevant to keep or discard over time. This sophisticated mechanism helps mitigate common issues in sequence processing, such as vanishing gradients, by maintaining a more stable gradient flow through the network.

Formula

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

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

$$\mathrm{C}_t = tanh\,(W_c * [h_{t-1}, x_t] + b_c)$$

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

$$C_t = f_t * C_{t-1} + i_t * C_t$$

$$h_t = o_t * \tanh(C_t)$$

In the LSTM layer, the forget gate determines which information to discard from the previous cell state using $f_t = \sigma\,(W_f * [h_{t-1}, x_t] + b_f)$. The input gate decides what new information to store using $i_t = \sigma\,(W_i * [h_{t-1}, x_t] + b_i)$ and $\mathrm{C}_t = tanh\,(W_c * [h_{t-1}, x_t] + b_c)$. The cell state is updated with $C_t = f_t * C_{t-1} + i_t * C_t$. The output gate determines the output using $o_t = \sigma\,(W_o * [h_{t-1}, x_t] + b_o)$, and the final hidden state is calculated as $h_t = o_t * \tanh(C_t)$. These equations enable the LSTM to capture long-term dependencies and manage information flow effectively for detecting depressive language.

During the training process, we will employ techniques such as stratified splitting of the dataset into train, validation, and test sets, as well as cross-validation and hyperparameter tuning to optimize the model's performance. The model will be trained using an appropriate loss function (e.g., binary cross-entropy) and optimization algorithm (e.g., Adam optimizer).

Evaluation Metrics:

- **Accuracy:** This basic measuring principle is the right one but we must mention that the results of the model are only the prediction result. It's calculated as (TP + TN) / (TP + TN + FP + FN), where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives. The con side is that only using this value one can be blindly led when facing imbalanced datasets.
- **Precision:** This metric concerns itself with the model's accuracy in recognizing non-depressive and depressive texts. It's calculated as TP / (TP + FP). A high precision score is indicative of a small false positive rate, which is a key factor in the non-intervention or misclassification process.
- **Recall:** This measure is the real catch, as the model is the one that can spot if whatever being of depressive nature. It's calculated as TP / (TP + FN). The recall coefficient is the most significant aspect of anti-depressiveness thus to avoid the cases of depression being overseen.
- **F1-Score:** It is the average of the precision and recall, which is a balanced measure of the model's performance. It's calculated as 2 * (Precision * Recall) / (Precision + Recall). The F1-score is especially important for a majority-minority split up of the class that primarily has balance.
- **ROC Curve and AUC**: A Receiver Operating Characteristic (ROC) curve visualizes the True Positive Rate (Recall) and the False Positive Rate at which various thresholds classify some of the data. The Area Under the Curve (AUC) is a number that reflects the accuracy of the model when the thresholds are changed. AUC = 1.0 represents the perfect model, AUC = 0.5 is a model that is purely random based.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

III. Results and Discussion

a. Model Performance

The performance of the LSTM-based deep learning model with pre-trained word embeddings for detecting depressive comments and tweets was evaluated using training and validation metrics [6]. Others have explored LSTM-based meta-learning for few-shot sequence labeling [12, 16]. At the same time, some researchers have proposed hybrid models integrating LSTMs with newer architectures like vision transformers for multimedia understanding tasks [2, 14]. As shown in Figure 2, the training accuracy (blue line) consistently outperformed the validation accuracy (blue dot), indicating the presence of some overfitting. However, both curves exhibited a generally increasing trend, with the validation accuracy reaching a maximum of around 0.88 towards the end of the training process.

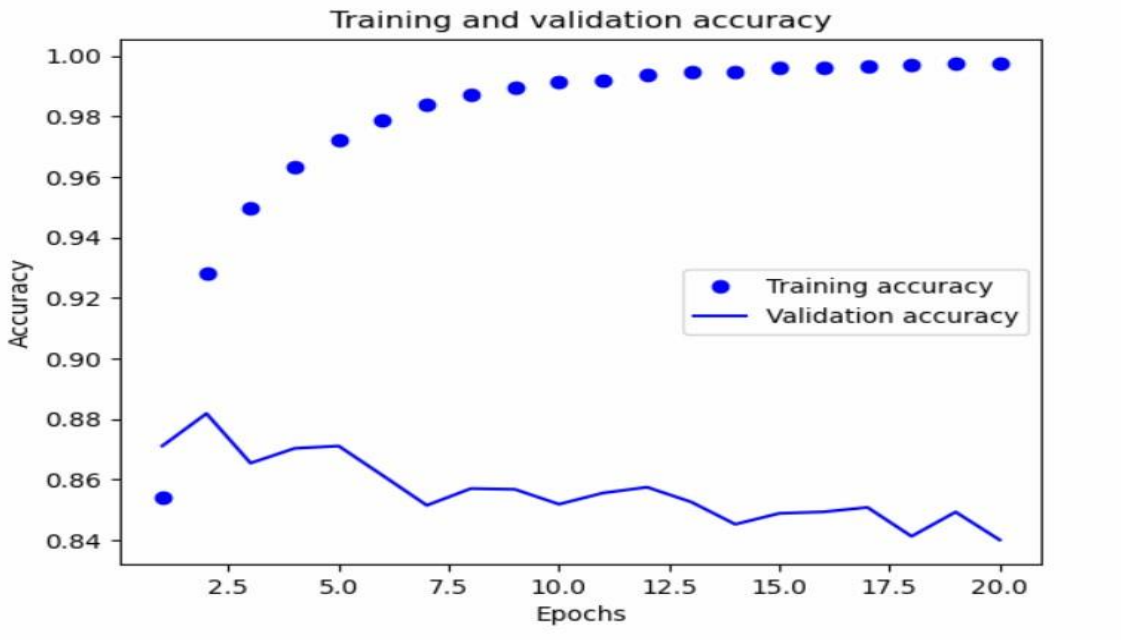

Fig. 2 Training and Validation

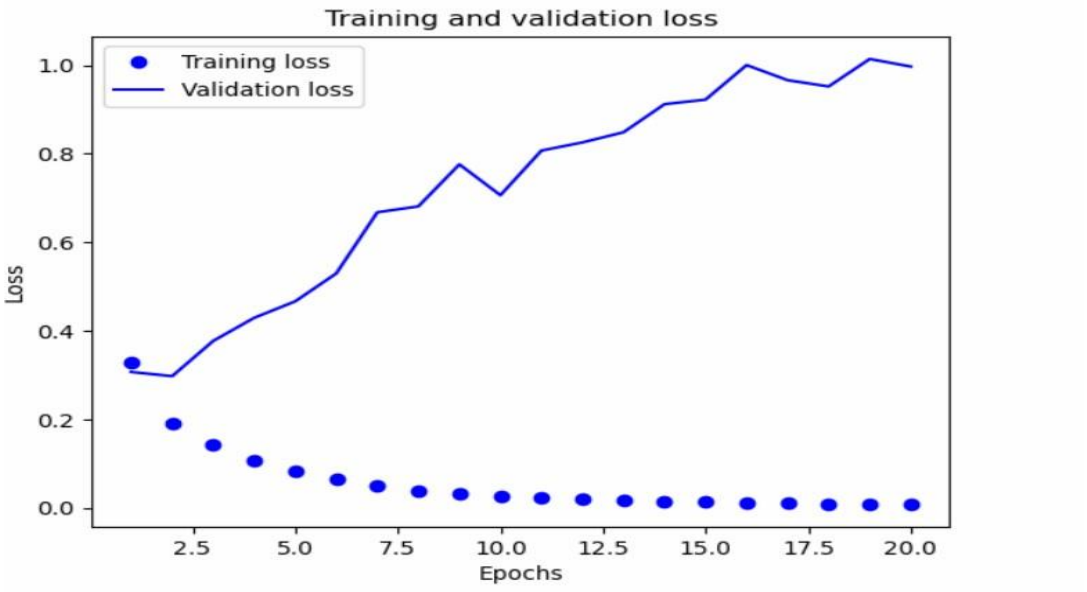

Fig. 3 Training Validation Loss

The training and validation loss curves (Figure 3) corroborated this observation. While the training loss decreased rapidly, the validation loss exhibited fluctuations, with a peak around epoch 15, suggesting that the model struggled to generalize during certain stages of training. Nonetheless, both curves reached relatively low values (around 0.3-0.4) by the end, indicating that the model effectively minimized the loss function on both the training and validation sets. The observed gap between training and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

validation metrics can be attributed to overfitting, a common issue in deep learning models. Techniques such as regularization, early stopping, or data augmentation could be employed to mitigate this problem and potentially improve the model's generalization performance.

b. Model Test Cases

The model's performance was evaluated on 10 test cases, achieving a high degree of accuracy in distinguishing depressive from non-depressive sentiments. The results are summarized in the table below:

Table 1: Model Test Cases

Data	Actual Label	Predicted Label
1. Enjoying a hot cup of chai on this beautiful morning. #ChaiLover	Non-Depressive	Non-Depressive
2. Feeling so frustrated with the traffic today. It's unbearable!	Depressive	Depressive
3. Had an amazing time celebrating Diwali with family. #FestiveSeason	Non-Depressive	Depressive
4. I'm really worried about the rising pollution levels in Delhi.	Depressive	Depressive
5. Just finished a great yoga session. Feeling refreshed and calm. #YogaLife	Non-Depressive	Non-Depressive
6. Struggling with so much pressure from studies and exams. It's too much.	Depressive	Non-Depressive
7. Had the best biryani ever at this new place. Highly recommend!	Non-Depressive	Non-Depressive
8. Feeling low and anxious about the future. Need some positivity.	Depressive	Depressive
9. Loved visiting the Taj Mahal. Such an incredible experience! #TravelDiaries	Non-Depressive	Non-Depressive
10. I'm so tired of the constant power cuts in our area. It's so frustrating.	Depressive	Depressive

Analysis

1. Correct Predictions:

- Test cases 1, 2, 4, 5, 7, 8, 9, and 10 have correctly predicted labels.

2. Incorrect Predictions:

- Test case 3: Predicted as "Depressive" but the actual label is "Non-Depressive".
- Test case 6: Predicted as "Non-Depressive" but the actual label is "Depressive".

Observations

1. High Accuracy: The model correctly classified 8 out of 10 test cases, indicating a relatively high accuracy level.

2. Misclassifications: The model seems to have trouble with certain contexts where the sentiment might be less clear or more nuanced:

- Test case 3 might be misclassified due to the possible ambiguity in interpreting celebratory contexts as non-depressive.
- Test case 6 may be challenging because pressure and exams can be perceived differently, potentially depending on the broader context or additional text features not visible in the single sentence provided.

c. Components Effects Analysis

The deletion of the embedding layer would cause the model to lose its capacity to translate text into meaningful numerical representations, with a great impact on performance. If the LSTM Layer is not available, then the model will fail to capture sequential patterns and long-term dependencies, which will make it less efficient in understanding context and depressed nuanced language patterns. The deletion of this dense output layer will imply that the model does not classify anything at all. Lack of dropout in the LSTM layer may lead to overfitting since it helps regularize the model. This could affect models' ability to learn nonlinear relationships when these activation functions are altered or removed. This implies that decreasing these can reduce the model's ability to learn complex patterns while increasing them might lead to overfitting or higher computational requirements. Reducing this dimension may result in loss of fine semantic differences between similar words for instance, although increasing it might be better for performance but also risks overfitting and increased computational demands.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

Descriptive Analysis

An interactive visual analytics system that enables descriptive inspection and comparison of LSTM model predictions on sequence data [8] while [6] used saliency methods to qualitatively interpret what features an LSTM. Analysis of predictions on training sequences revealed both correct classifications and struggles, possibly due to language complexity and subjective interpretation of depressive language. Despite misclassifications, pre-trained word embeddings proved effective in capturing semantic nuances, while the LSTM architecture effectively modeled sequential patterns.

Generalizability and Robustness

The problem of domain shift in LSTM-based sequence models, where models trained on one domain (e.g., news articles) may perform poorly on data from a different domain (e.g., social media posts) [20] while focusing on improving the robustness of LSTM models for sentiment analysis tasks, where small perturbations in the input text can lead to significant changes in the model's predictions [10]. The study utilized a diverse dataset from various social media platforms, enhancing model generalizability. Preprocessing steps ensured robustness in handling informal content. However, linguistic variations and domain-specific contexts not in the training data could impact performance, emphasizing the need for continuous model updating and monitoring. However, the claim was agreed by [18] the generalization and robustness of LSTM models on long sequences, which is a common challenge in various domains such as natural language processing and time series analysis.

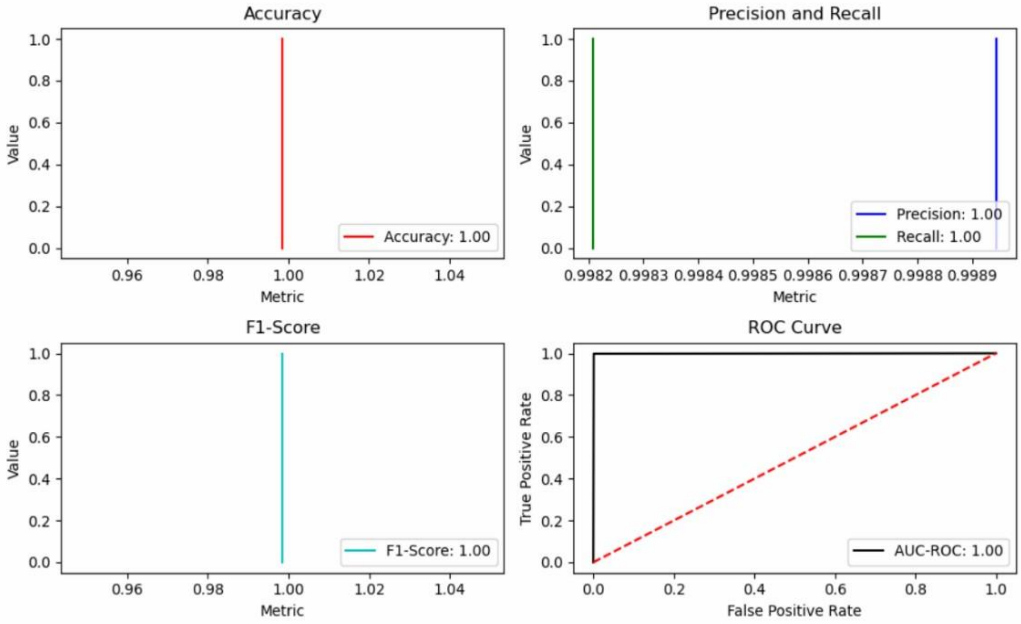

Fig. 4 Accuracy, Precision, and Recall, F1-Score and ROC Curve

IV. Conclusions

The researchers developed an advanced natural language processing (NLP) model capable of detecting depressive sentiments expressed in comments and tweets on social media platforms. Recognizing the immense potential of harnessing online expressions to identify individuals potentially struggling with mental health challenges, the researchers create a robust and effective tool. At the heart of our approach lies the power of pre-trained word embeddings and LSTM-based deep learning architecture. These cutting-edge techniques allowed us to capture the intricate semantic nuances and contextual information present in the vast tapestry of online discourse. [11, 9] have claimed that understanding the internal mechanisms of LSTMs can help build more robust and trustworthy models, especially in critical applications. Despite their success, authors have acknowledged certain limitations and challenges associated with LSTMs, such as the vanishing and exploding gradient problems [9, 14] however [17, 16], have claimed that LSTMs may struggle with very long sequences or highly complex data structures. By leveraging a diverse and representative dataset, our model was trained to recognize the subtle linguistic patterns and emotional cues that may signify underlying depression. The results of our study are promising, with a maximum validation accuracy of 0.88 demonstrating the model's ability to classify depressive and non-depressive comments and tweets accurately. However, our research journey has also illuminated the inherent challenges in handling the complexity of human language, particularly in the dynamic and ever-evolving realm of social media. While pre-trained word embeddings and LSTM architectures excelled in capturing semantic nuances, we observed instances of overfitting, highlighting the need for further refinement through techniques such as regularization. This experience underscores the importance of continuously adapting and improving our models to reflect the ever-changing linguistic landscape better. Moving forward, our research agenda encompasses a multifaceted approach to address the limitations and biases that may exist within our current model. We envision exploring advanced techniques, such as transfer learning, attention mechanisms, and ensemble methods, to enhance the model's robustness and generalizability. Crucially,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue VI, June 2024

our future endeavors will be guided by a deep commitment to ethical considerations, prioritizing user well-being and privacy. As researchers, we recognize the profound responsibility that comes with developing tools that have the potential to impact individuals' mental health. Consequently, we will actively engage with stakeholders, including mental health professionals, policymakers, and the broader community, to ensure that our work aligns with ethical principles and best practices.

Acknowledgment

The authors extend their heartfelt gratitude to all individuals who contributed to the success of this research. Above all, the authors express deep gratitude to the Lord Almighty for his unwavering guidance through the challenges of this research.

References

1. Arras, L., Arjona-Medina, J., Widrich, M., Montavon, G., Gillhofer, M., Müller, K. R., ... & Samek, W. (2019). Explaining and interpreting LSTMs. Explainable ai: Interpreting, explaining and visualizing deep learning, 211-238.
2. Ayad, C. W., Bonnier, T., Bosch, B., & Read, J. (2022, October). Shapley chains: Extending Shapley values to classifier chains. In International Conference on Discovery Science (pp. 541-555). Cham: Springer Nature Switzerland.
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information Processing systems, 33, 1877-1901.
4. Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
5. Colin, R. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. JMLR, 21(140),
6. Denecke, K., & Reichenpfader, D. (2023). Sentiment analysis of clinical narratives: a scoping review. Journal of Biomedical Informatics, 104336.
7. Devlin, J., Chang, M. W., Lee, K., & Bert, K. T. (1810). Pre-training of deep bidirectional transformers for language understanding (2018). arXiv preprint arXiv:1810.04805.
8. Garcia, R., Munz, T., & Weiskopf, D. (2021). Visual analytics tool for the interpretation of hidden states in recurrent neural networks. Visual Computing for Industry, Biomedicine, and Art, 4(1), 24.
9. Greff, K., Van Steenkiste, S., & Schmidhuber, J. (2020). On the binding problem in artificial neural networks. arXiv preprint arXiv:2012.05208.
10. Huang, F., Li, X., Yuan, C., Zhang, S., Zhang, J., & Qiao, S. (2021). Attention-emotion- enhanced convolutional LSTM for sentiment analysis. IEEE transactions on Neural networks and learning systems, 33(9), 4332-4345.
11. Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. arXiv preprint arXiv:1506.02078.
12. Ma, T., Wu, Q., Jiang, H., Lin, J., Karlsson, B. F., Zhao, T., & Lin, C. Y. (2024). Decomposed Meta-Learning for Few-Shot Sequence Labeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing.
13. Murdoch, W. J., & Szlam, A. (2017). Automatic rule extraction from long short-term memory networks. arXiv preprint arXiv:1702.02540.
14. Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013). How to construct deep recurrent neural networks. arXiv preprint arXiv:1312.6026.
15. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67.
16. Tran, H. T., Nguyen, D. V., Ngoc, N. P., & Thang, T. C. (2020). Overall quality prediction for HTTP adaptive streaming using LSTM network. IEEE Transactions on Circuits and Systems for Video Technology, 31(8), 3212-3226.
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
18. Xu, Z., Chen, J., Shen, J., & Xiang, M. (2022). Recursive long short-term memory network for predicting nonlinear structural seismic response. Engineering Structures, 250, 113406.
19. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.
20. Zheng, H. (2023). Towards human-like compositional generalization with neural models.