Depression Detection using Deep Learning Algorithms

Alaa Zaghloul
Department of Computer Science
Misr University for Science and Technology
Cairo, Egypt
alaa.zaghloul@must.edu.eg

Omar Khaled
Department of Artificial Intelligence
Misr University for Science and Technology
Cairo, Egypt
94105@must.edu.eg

Rania Elgohary
Department of Artificial Intelligence
Misr University for Science and Technology
Cairo, Egypt
rania.elgohary@must.edu.eg

***Abstract:* This research study aims to provide a depression detection project that uses text analysis and natural language processing (NLP) to identify symptoms of depression. In order to conduct sentiment analysis on big datasets of tweets, this project will employ a deep learning model. Social media platforms have evolved into places where individuals express their ideas and feelings. Our objective is to create a chat platform that enables users to interact with friends, coworkers, or complete strangers while using text analysis to identify sadness. There are several browsers that can be used to visit the website and guidance on interacting with it. The significance of early depression detection and its possible effects on community well-being—including detrimental effects on local company productivity and healthcare costs—will be emphasized in our research. The purpose of this project is to increase public awareness of the advantages of early identification and to offer a deep learning-based approach to assist people in identifying depression and obtaining the necessary assistance.**

Keywords: Natural Language Processing, Sentiment Analysis, Deep learning, Depression detection.

I. INTRODUCTION

Depression is a severe mental illness that affects a large number of people worldwide. It causes emotional suffering, a decrease in interest in activities, and major behavioral changes.[1] Major depressive disorder is diagnosed based on the patient's self-reported experiences, a mental state assessment, and the opinion of family members or friends regarding the patient's conduct. It may be difficult to diagnose the illness because there are no laboratory testing for it.[2] But with more people interacting on social media, it's now feasible to identify depressive signs early on, even before doing more conventional diagnostic tests. In this research work, we describe a deep learning and natural language processing (NLP) approach for depression identification by tracking users' tweets on Twitter.[3] Our goal is to create a technology that can identify depression from data from Twitter tweets, allowing those who are depressed to receive early assistance. We introduce a conversation website with a deep learning system that analyzes text to assess if a user is depressed and offers suggestions for treatment.[1] Early detection of depression will hopefully lessen its impact on both individuals and communities, improving the quality of life for those who are impacted. Technological developments and the increasing usage of social media and the internet have created new avenues for the detection and tracking of mental health issues in recent years. By using a deep learning algorithm and a website to analyze text data, this research seeks to advance the science of depression identification by offering early intervention to depressed individuals. The website provides a convenient and approachable medium for people to publish textual content, such as blog entries, social media updates, or personal essays. These textual inputs provide insightful information about a person's feelings, ideas, and behavioral patterns, all of which can be markers of their mental health. Deep learning algorithms and natural language processing (NLP) can be used to find hidden patterns and signals that could point to the presence of depression. [4].

The purpose of this study is to create a deep learning algorithm and website for the purpose of developing a depression detection system.[8] After analyzing the input text, the deep learning algorithm will estimate whether or not the person exhibits depressive symptoms.[7] There are multiple crucial steps in the process. First, after pre-processing a dataset containing text samples from people with and without depression, the text data will be tokenized, lemmatized, and converted into numerical representations that can be fed into deep learning algorithms [7]. The design of the deep learning architecture and natural language processing (NLP) will be based on recurrent neural networks (RNNs), more precisely Long Short-Term Memory (LSTM) units [5]. LSTM units are perfect for text data analysis since they are well-suited for modeling sequential data. The tagged dataset will be used to train the deep learning model, which will aim to identify patterns and characteristics that differentiate between text samples that are depressed and those that are not [7].

The model will be implemented on the website to give real-time depression detection when it has been trained. Through an intuitive interface, users will be able to enter text data, and the models will produce predictions on their depression status. Users can utilize the forecasts to better understand their mental health and, if needed, to be reminded to seek professional assistance. We also made some comparisons between various performers' performances. machine learning techniques to determine which will function best with the chosen dataset [9].

The ultimate goal of this research is to create a novel depression detection system by utilizing the capabilities of a deep learning algorithm and a website. This system has the ability to provide early intervention and support for those who are experiencing depression by analyzing text data using deep learning models and natural language processing techniques. In the end, technology and mental health care combined have the power to completely transform the way depression is identified and treated, improving the quality of life and general wellbeing of those who suffer from this difficult illness.[2]

The remnant of the paper is organized as follows. Related Work is explained in section 2, dataset is explained in section 3, proposed Algorithms is explained in section 4, experimental results are presented and discussion in section 5, references will be in section 6.

II. Related Works

Prior to Large user-generated data sets are typically the foundation for text-based depression detection on sparse data [5]. Less research has been done on clinical discussions with sparse scripts. This study suggests using pretrained word embeddings in a text-based multi-task BGRU network to simulate clinical interview responses from patients. Our primary method points to modeling both depression riskiness and binary health state using a unique multi-task loss function. We independently examine the use of large-data pretraining for depression identification, as well as word- and sentence-level word embeddings. We present mean averaged results from numerous independent runs on dispersed data to bolster our conclusions. First, we demonstrate the supplementary role of pretraining in word-level text-based depression identification. Second, our findings clarify that word-level embeddings should generally be avoided in favor of sentence-level embeddings. Mean and attention gathering should be prioritized over last-timestep gathering, even though the choice of gathering function is less important. Our approach culminates in a macro F1 score of 0.84 and MAE of 3.48 on the DAIC-WOZ development set, along with results on the existence of depression and the predicted severity score.

A novel multi-task model architecture that integrates severity prediction and binary depression detection. the distinction in performance between sentence-level and word-level embedding.

The need to develop intelligent systems that can effectively handle early risk detection (ERD) issues on social media stands, such as early depression detection, early hearsay detection, or identification of numerous sexual predators, is highlighted in A Text Classification Frame for Simple and Effective Early Depression Detection Over Social Media Stands. These systems, which these days are primarily built using machine learning approaches, need to be able to handle data flow because users contribute data over time. These systems also need to be able to judge when the amount of processed data is adequate for user classification. Furthermore, such systems must be able to defend their choices because ERD tasks require important choices that may have an impact on people's lives [10]. However, most standard, and state-of-the-art supervised machine learning models, (such as SVM, MNB, Neural Networks, etc.) are not well suited to deal with this screenplay.

concentrate on early depression detection as an important ERD task. Depression is described as a serious intellectual disorder that affects people's ability to communicate, particularly through their emotions, and, supposedly, how they react with others in Multi-Task Learning for Depression Detection in Dialogs. This study looks at depression signs in dialogs, a less researched environment where social media data flow is problematic. We suggest utilizing topic and dialog act prediction to investigate the influence of dialog structure in light of our hypothesis that depression and emotion can influence one another [3].

Psychological analysis examines a large amount of text and extracts key information, characteristics, and facts from users' points of view in Psychological Analysis for Depression Detection from Social Networking Sites [1]. Social networks are used by psychological analysts to identify behaviors and activities associated with depression. Social networks offer a wealth of information on the mindsets and behaviors that precede the beginning of depression, including low social, seeking medical attention, placing a high value on oneself, and engaging in high levels of daytime and evening activity. multiplicity by the use of several algorithms and modules with a large amount of data.

Detecting depression from social media platform data with machine learning methods Social media platforms have emerged as an excellent means for individuals to interact with their watchful friends and exchange thoughts, images, and videos that capture their feelings, senses, and moods [2]. This makes it possible to examine user moods and attitudes when they communicate using these online tools by analyzing social media platform data for users' feelings and emotions. Facebook users' high fineness answers to intellectual health concerns are identified by machine learning approaches.

Abbreviations and Acronyms:

Abbreviation	**Definition**
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
NLP	Natural Language Processing
MAE	Mean Absolute Error
BGRU	Bidirectional Gated Recurrent Unit
AUC-ROC	Area Under the Curve of the Receiver Operating Characteristic curve.
SVM	Support Vector Machine
MNB	Multinomial Naïve Bayes
ERD	Early depression detection

III. Dataset

The dataset is collected by web scraping method on Twitter to collect tweets and store them, the dataset contains three features: tweet_ID, tweet, and the label. tweet_ID feature has been dropped while preprocessing the data, the tweet feature contains a wide range of tweets and they were pre-processed and transformed into numerical values to prepare them as an input to the LSTM algorithm, the label feature contains binary values as zeros means that the person is not depressed, and ones means that he is depressed.

IV. METHODOLOGY

The proposed approach uses deep learning techniques to identify references to depression. Preprocessing the data entails eliminating unnecessary information and retaining only that which is necessary.

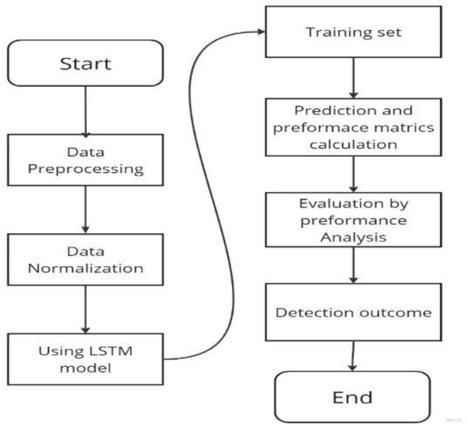

Figure 1. The flowchart of the Proposed algorithm

A. Used Algorithms:

In this Research, we implemented multiple Deep learning and Machine Learning models to compare their ability in detecting the depression of a human based on input text, we used LSTM, MNB, Logistic regression and SVC algorithms.

- Long Short-Term Memory:

Long Short-Term Memory networks, or LSTMs for short, were created to solve a major issue that traditional Recurrent Neural Networks (RNNs) faced: the "long-term dependency problem." In theory[7], RNNs might potentially connect historical data to the work at hand, but they struggle when there is a significant gap between relevant data and its real-world implementation. LSTMs, a specialized version of RNNs, provide a way around this problem. Unlike conventional RNNs, which have simple repeating modules, LSTMs have four interacting layers that are specifically engineered to capture and leverage long-term dependencies. The cell state, a horizontal line acting as an information conveyor belt, is at the core of the LSTM idea. LSTMs use gates to selectively allow or prohibit information flow across the cell state, such as the input gate and forget gate layers. LSTMs are very good at learning and remembering long-term dependencies because they use a methodical walkthrough to decide what data to update, forget, and output. Because of this unique architecture, LSTMs have shown themselves to be an effective and popular tool in the field of deep learning, with notable success observed in a wide range of applications.

Below are the LSTM input outputs and the corresponding equations for a single timestep.

1. Forget gate formula:

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

2. Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

3. Cell Update:

$$\tilde{C}_t = \tanh\,(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

4. New cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

5. Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

6. New hidden state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

Where:

- f_t: forget gate vector.
- i_t: input gate vector.
- $\tilde{C}_t$: cell update vector.
- C_t: cell state vector.
- o_t: output gate vector.
- h_t: hidden state vector.
- x_t: input vector.
- W_f,W_i,W_C,W_o: weight matrices for forget, input, cell update, and output gates, respectively.
- b_f,b_i,b_C,b_o: bias vectors for forget, input, cell update, and output gates, respectively.
- σ: sigmoid function.
- tanh: hyperbolic tangent function.
- $*$: element-wise multiplication.

- Multinomial Naïve Bayes:

Naïve bayes is a probabilistic technique used to build data categorization models. The idea of probability—more precisely, the possibility that data belongs to a specific class—is central to this strategy. From Naïve Bayes, two main models are derived: Gaussian and Multinomial. The Multinomial model performs exceptionally well in the classification of non-numerical data, exhibiting notably lower complexity and the capacity to carry out classification with short training sets without requiring ongoing retraining.

Bayesian Probability Formulae:

$$P(c \mid d) = \frac{P(c) \cdot P(d|c)}{P(d)} \quad (7)$$

Where:

- $P(c \mid d)$ is the posterior probability of class (c, target) given predictor (d, attributes).
- $P(c)$ is the prior probability of class.
- $P(d \mid c)$ is the likelihood which is the probability of predictor given class.
- $P(d)$ is the prior probability of predictor.

In Multinomial Naive Bayes, the likelihood $P(d \mid c)$ is calculated as follows:

$$P(d \mid c) = P(w_1, w_2, \ldots, w_n \mid c) \quad (8)$$

$$P(w_1 \mid c) \cdot P(w_2 \mid c) \cdot \ldots \cdot P(w_n \mid c) \quad (9)$$

Where:

- $w_1, w_2, \ldots, w_n$ are the words in document d.
- $P(w_1 \mid c)$ is the probability of word w_1 appearing in a document of class c.

- **Logistic Regression:**

Logistic regression is an effective tool for binary classification applications. Logistic regression, in contrast to its linear equivalent, is intended to forecast the likelihood that an instance is a member of a specific class, usually assigning values of 0 or 1. Logistic regression can be described as a flexible and extensively used technique, renowned for its ease of interpretation and simplicity. The logistic function combines the coefficients that the algorithm predicts for each attribute in a linear fashion to yield probability. For this reason, logistic regression is particularly useful when analyzing the relationship between independent factors and the likelihood of a binary outcome.

Logistic Regression Formulae is expressed as:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad (10)$$

Where:

- $P(Y = 1 \mid X)$ is the probability that the binary output variable Y is 1 given the input variables X.
- β_0 and β_1 are the parameters of the model.
- e is the base of the natural logarithm, approximately equal to 2.71828.

- **Support Vector Machine :**

Support Vector Machine (SVM) is a highly efficient supervised machine learning technique that may be applied to both regression and classification tasks. SVM creates a hyperplane, or optimal decision border, by dividing data points by use the maximum margin. It offers the Support Vector Classifier (SVC), Support Vector Machine (SVM), and Maximal Margin Classifier (the MMC) for nonlinear data using a kernel approach. Along the road, SVM has come across kernels like Sigmoid, Polynomial SVM, and Radial Basis Function that are intended for different situations. Understanding SVM requires an understanding of key factors such as regularization (C), gamma, kernel selection, and degree of flexibility. In the end, SVM has drawbacks like susceptibility to outliers but also advantages like handling nonlinear data and being effective in high-dimensional areas.

SVC formula can be expressed as follows:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) \quad (11)$$

Where:

- β_0 is the bias term.
- n is the number of support vectors.
- $\alpha_i$are the dual coefficients.
- y_i are the class labels of the support vectors.
- x_i are the support vectors.
- $K(x_i, x)$ is the kernel function.

The predicted class label for x is then obtained by taking the sign of $f(x)$:

$$\hat{y} = sign(f(x)) \quad (12)$$

B. Text processing: Regular expressions are used to specify and extract hashtag, username, and URL patterns from the text. Regular expressions are also used to identify and remove emojis from text. The contraction dictionary is used to expand contractions (e.g., 'don't' to 'do not'). To separate words, spaces are placed around slashes and non-alphanumeric characters and symbols are eliminated.

The training and testing datasets are split into two parts in the next phase: 80% for training and 20% for testing. Applying the algorithm and testing the suggested model are the final steps in the process.

C. Sequential Model: is a type of neural network architecture used in deep learning. It is called "sequential" because the network is made up of a sequence of layers, where the output of one layer is passed as input to the next layer. In other words, the data flows through the layers in a sequential.

D. Embedding Layer: A neural network's embedding layer is a particular kind of layer that aims to learn how to identify more significant patterns in the input data, potentially improving performance. It will therefore obtain embedding for every sentence in the dataset. The job complexity and the vocabulary's girth must be taken into consideration while determining the hyperparameter's girth. Afterwards, further layers in the neural network might receive the output of the embedding layer.

E. Dropout Layer: A typical problem known as overfitting happens when a deep learning model learns to match training data too well, and the model performs poorly on new, unseen data. The dropout rate is the total number of neurons in a layer that are eliminated at random during training.

F. LSTM Layer: LSTM is a type of Recurrent Neural network; they are capable of learning long-term dependencies in data. This is because they use a memory cell to store information from previous inputs, allowing them to remember the context of the data and use it to make predictions. Additionally, LSTM networks use gates to control the flow of information, allowing them to selectively choose which information to keep and which to discard.

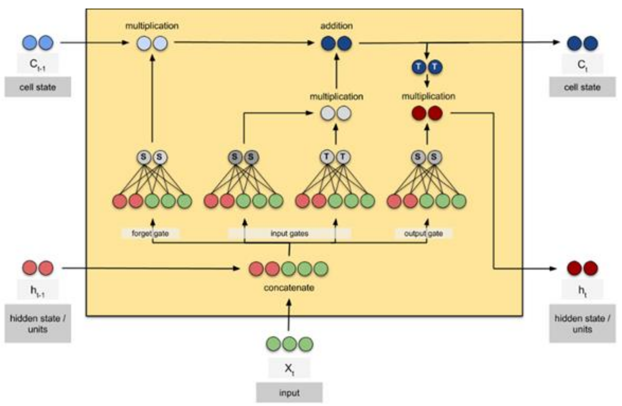

Figure 2. The structure in the LSTM algorithm

G. Dense Layer:
This layer, which helps connect all the layers to produce correct output and a fully connected layer, is always exercised toward the end of the model. All layer's neurons extract input from the layer before it in a way that connects every layer to every other layer. Whether the output is binary or consists of several outputs, its shape is defined.

H. Training and evaluation:
Using the fit function and a predetermined batch size and number of epochs, the model is trained. The prediction function is used to make predictions on the test set. Predicted probabilities are converted into binary predictors using a threshold of 0.5, 0.3, 0.2 Evaluation measures, including accuracy, recall, F1 score, precision and ROC-AUC score, are calculated using functions from scikit-learn library in python. Confusion matrix is calculated using real labels and predicted binary labels and visualized using a heatmap. Rating metrics are stored in a data frame for further analysis or reporting.

V. Discussion

After implementing the LSTM and the other Machine learning algorithms, it is necessary to evaluate each one of them and compare their results to see which algorithm performed better, as shown in table 3 the LSTM algorithm was the best algorithm among all the others, it showed high accuracy in training and validation. In figure 3, we visualized the performance of the model in another way using the confusion matrix to see it’s ability in classifying the depressed from non-depressed tweets, the algorithm showed high ability in classifying the tweets without overfitting. Then we compared the LSTM model with Adam and SGD as optimizers while maintaining the same hyperparameters to see which optimizer is the best, as shown in figure 11, Adam optimizer is way much better than SGD.

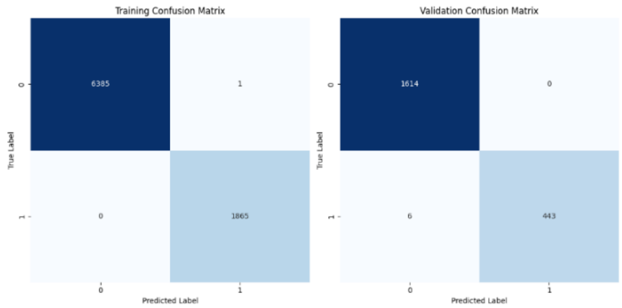

Figure 3. Confusion matrix of training and validation at epoch 10

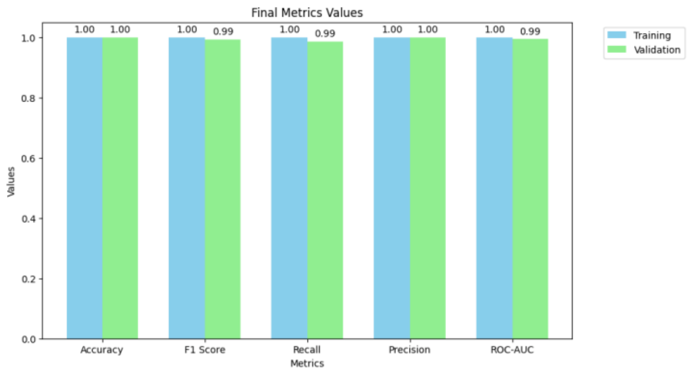

Figure 4. The performance LSTM

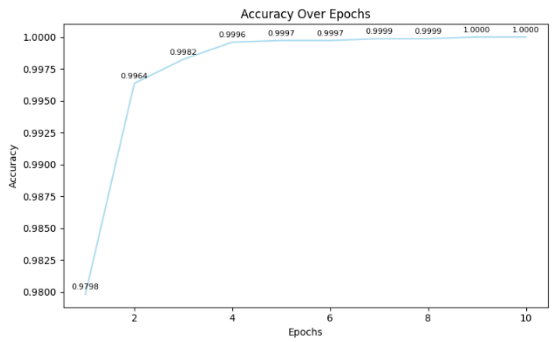

Figure 5. Accuracy of LSTM over epochs

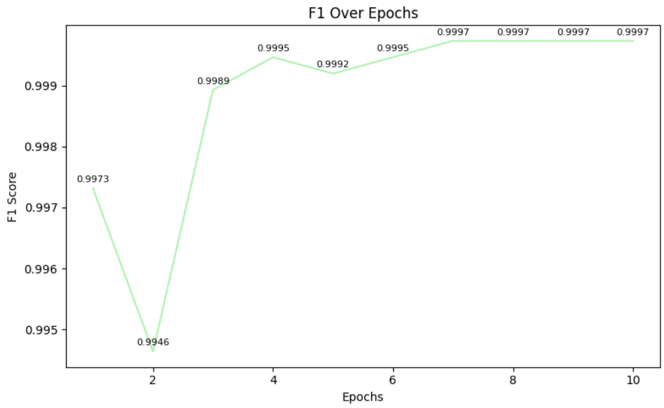

Figure 6: F1-score of LSTM over epochs

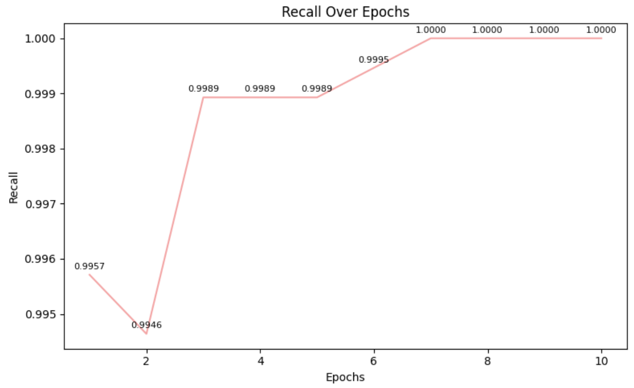

Figure 7: Recall values of LSTM over epochs

Table 2. Parameter setting

Parameters	Definitions	Values
E	Epoch	10
lr	Learning rate	0.01
D	Dropout	0.3

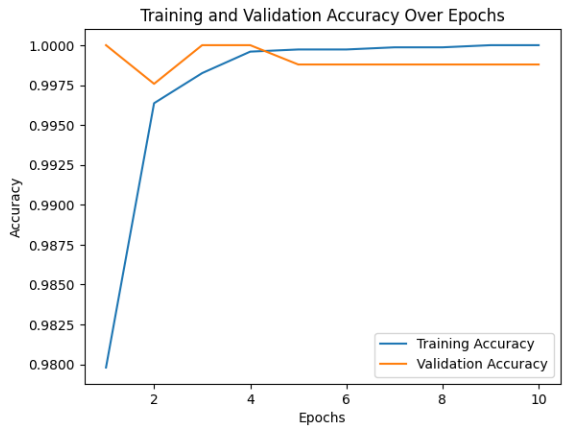

Figure 8. The training and validation Accuracy of the proposed algorithm at Epoch 10

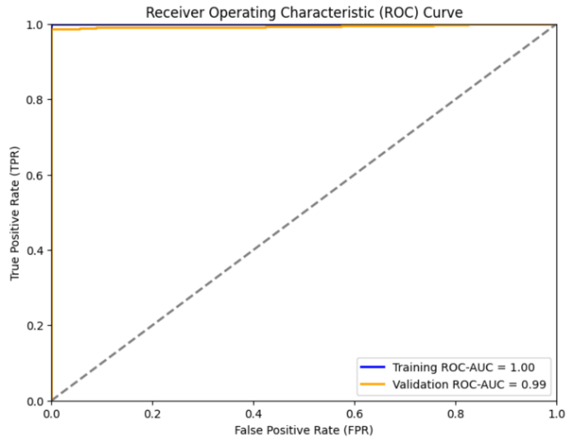

Figure 9: ROC-AUC curve of LSTM with Adam optimizer

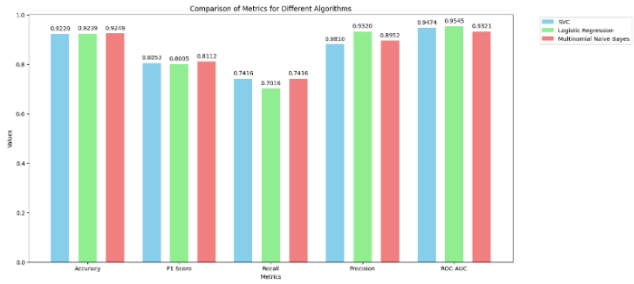

Figure 10. Comparison between different metrics of all machine learning algorithms

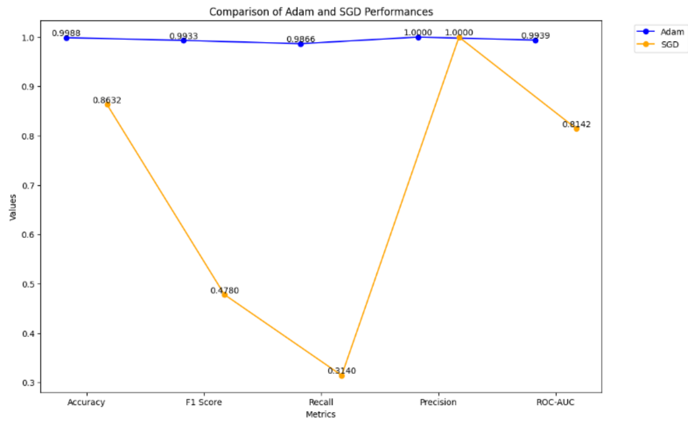

Figure 11. The performance of SGD and ADAM optimizers on LSTM

Table 3. comparison between the proposed algorithm and other algorithms

Model	Accuracy	Precision	Recall	F1-Score	AUC
LSTM	99.8%	100%	98.6%	99.3%	99.3%
SVC	92%	88.1%	74.%	80%	94.7%
Logistic regression	92.3%	93.2%	70.1%	80%	95.4%
MNB	92.4%	89.5%	74.1%	81.1%	93.2%

VI. CONCLUSION

This research provided an Artificial Intelligence approach to detect depression using tweets data to provide early diagnosis about the mental health of human. The analysis showed that LSTM algorithm was the best algorithm to use on the dataset achieving 99.8% a

ACKNOWLEDGMENT

We would like to sincerely thank everyone who helped us to successfully complete this research project on the use of deep learning and natural language processing (NLP) in the diagnosis of depression.

REFERENCES

[1] C. Intelligence and Neuroscience, “Retracted: Psychological Analysis for Depression Detection from Social Networking Sites,” Comput. Intell. Neurosci., vol. 2023, pp. 1–1, 2023, doi: 10.1155/2023/9796187.

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

[3] C. Li, C. Braud, and M. Amblard, “Multi-Task Learning for Depression Detection in Dialogs,” 2022, [Online]. Available: http://arxiv.org/abs/2208.10250

[4] R. Kaur, A. Majumdar, P. Sharma, and B. Tiple, “Analysis of Tweets with Emoticons for Sentiment Detection Using Classification Techniques BT - Distributed Computing and Intelligent Technology,” A. R. Molla, G. Sharma, P. Kumar, and S. Rawat, Eds., Cham: Springer Nature Switzerland, 2023, pp. 208–223.

[5] H. Dinkel, M. Wu, and K. Yu, “Text-based depression detection on sparse data,” pp. 1–12, 2019, [Online]. Available: http://arxiv.org/abs/1904.05154

[6] S. G. Burdisso, M. Errecalde, and M. Montes-Y-Gómez, “UNSL at Erisk 2019: A Unified Approach for Anorexia, Self-harm and Depression Detection in Social Media,” CEUR Workshop Proc., vol. 2380, pp. 9–12, 2019.

[7] K. S. Nugroho, I. Akbar, A. N. Suksmawati, and Istiadi, “Deteksi Depresi Dan Kecemasan Pengguna Twitter Menggunakan Bidirectional Lstm,” arXiv, no. Ciastech, pp. 287–296, 2023, doi: 10.48550/arXiv.2301.04521.

[8] S. Gupta, A. Singh, and J. Ranjan, “Multimodal, multiview and multitasking depression detection framework endorsed with auxiliary sentiment polarity and emotion detection,” Int. J. Syst. Assur. Eng. Manag., vol. 14, no. 1, pp. 337–352, 2023, doi: 10.1007/s13198-023-01861-z.

[9] S. Jickson, V. S. Anoop, and S. Asharaf, “Machine Learning Approaches for Detecting Signs of Depression from Social Media BT - Proceedings of International Conference on Information Technology and Applications,” S. Anwar, A. Ullah, Á. Rocha, and M. J. Sousa, Eds., Singapore: Springer Nature Singapore, 2023, pp. 201–214.

[10] Burdisso, Sergio G., Marcelo Errecalde, and Manuel Montes-y-Gómez. "A text classification framework for simple and effective early depression detection over social media streams." Expert Systems with Applications 133 (2019): 182-197.