Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 1s (2025)
https://www.jneonatalsurg.com
OPEN ACCESS

Social Media-Driven Depression Detection Using Improved Recurrent Neural Architecture Leveraging GloVe Word Embeddings

Tariq Siddiqui[*1], Dr. Ashish Pandey[2]

[*1]Research Scholar, CSE, Bhabha University, Bhopal, M.P. 462001, India.

[2]Associate professor, CSE, Bhabha University, Bhopal, M.P. 462001, India.

Email ID: ashishbpl88@gmail.com

*** Corresponding Author:**

Tariq Siddiqui

Email ID: tariq.qazi007@gmail.com

Cite this paper as: Tariq Siddiqui, Dr. Ashish Pandey, (2025) Social Media-Driven Depression Detection Using Improved Recurrent Neural Architecture Leveraging GloVe Word Embeddings. *Journal of Neonatal Surgery*, 14 (1s), 1392-1399.

ABSTRACT

The increasing incidence of mental health disorders, especially depression in modern society highlights the immediate need for effective methods to identify and reduce these issues immediately. This study introduces an advanced recurrent nervous network (RNN) model to classify depression based on lesson data obtained from social media platforms, which has greatly improved by incorporating the gloves (global archives for word representation) embeding. The use of gloves embeding provides the model enhanced semantics and relevant insights, which facilitates more intensive understanding of micro-linguistic elements found in social media communication related to depression. The model trained and evaluated using reddit dataset. Experimental conclusions revealed exceptional display matrix, with a test accuracy of 97.22%. These results display the proficiency of the model in correcting manifestations related to depression while maintaining both high sensitivity and uniqueness. By merging linguistic understanding with machine learning, the proposed structure presents a viable result for the initial identity of depression through public platform media text analysis

Keywords: *Depression Classification, RNN , Social Media Text , Glove Embedding , Reddit Dataset , Mental Health*

1. INTRODUCTION

Depression is recognized as a concern for a pressure psychological good, which affects an important section of global people [1]. Current research estimates that about 20% of people will experience depression at some point in their lives. In addition, a striking 4.5% of the global population is believed to be currently living with this condition. Unfortunately, depression is often undesirable and untreated in many areas, resulting in harmful effects on self -esteem and, in extreme cases, suicide has an elevated risk. Therefore, timely identification and effective treatment are paramount to effectively manage depression and increase the overall welfare of the affected people. In the current decade, online social-media platforms are increasing as a means of identifying potential indicators of depression.

A significant growth in the field of natural language processing (NLP) is a long -term short -term memory (LSTM) model, which provides reliable technology to evaluate and understand sequential data, including lessons [2]. In therapeutic settings, the application of deep teaching models such as LSTM can help detect depression and guarantee that people receive timely treatment and relevant referrals. By using these models, professionals can keep an eye on signs of depression on social media and provide active support by providing resources and advice that are particularly corresponding to the individual. Cooperation of technology experts and mental health professionals is necessary as it enables the integration of state -of -the -art technical solutions with clinical knowledge to promote early identification and efficient support systems.

Additionally, social media and online forum can act as encouraging spaces, who can give to people who are struggling with depression, which is a sense of room dadri and equality [3]. A label classification dataset, which includes samples of the text of a variety of social media sites, which is classified as a thoughtful of feelings of depression or non-disdain, used to compare multiple LSTM variations in this work goes. Assessing the effectiveness of many LSTM models in identifying depression in lessons obtained from social media sites is the main goal of this investigation.

This letter introduced the novel technique for depression classification for social-media text with existing methods comparatively with gloves embeding with gloves embeding, and the proposed technology verified with high accuracy than

2. RELATED WORK

Significant progress has been made in the field of natural language processing (NLP), especially with the use of long short - term memory (LSTM) network. This scholarly examines check the use of many LSTM models and how well they handle and analyze text data in different fields. Due to their adaptability, LSTM architecture text classification, emotion analysis and machine translations are becoming the necessary tools to improve the efficiency of NLP applications.

The effectiveness of the LSTM model using Twitter data was quoted as [4] in a paper, which focuses on the difference between a balanced and unbalanced dataset. While the unbalanced dataset was subjected to an oversampling technique to reduce the impact of missing values, a balanced dataset was divided into positive and negative emotions by researchers. Many machine learning techniques, such as support vector machines (SVMs), Decision Tree, K-Nearest Neighbor (KNN), and LSTM, were used in evaluation. With an outstanding accuracy rate of 83% in both dataset types, the results showed that LSTM improved others, exposing its ability to emotion analysis tasks.

Another work [5] focused on creating a bespoke dataset from raw Twitter data, thus expanding the use of LSTM in social media environment. Unnecessary data was converted into examples labeled by preprocessing procedures, making them suitable for intensive teaching model training. The recruitment nerve network (RNN), Convolution Neural Network (CNN), and Gaped Recruitment Units (GRU) were tested between algorithms used. The GRU model showed a remarkable accuracy of 98%, indicating its efficacy in text classification functions.

To address mental health issues, the research reported in [6] used natural language processing (NLP) techniques to analyze reddit data and find a tendency to find a tendency of depression in user-related materials. The study used classifier including SVM, Logistics Regression, Random Forest, and Multi-Lear Perceptron (MLP), which included various types of feature extraction techniques such as N-Gram and linguistic inquiry and Word Count (LIWC). With the best accuracy of 91%, a combination of LiwC, LDA, and N-Grams, processed via MLP, demonstrated the promise of lesson analysis in identifying mental health disorders.

Another study [7] focused the focus of analysis by combining text analysis with audio classification using various models for each modality. Researchers insist on detection of depression, with F1-score of 0.8 for text data and a score of 0.75 for audio analysis, which is related to voice quality using CNN for CNN and sequence modeling using CNN and sequence modeling For LSTM. The importance of multimodal analysis in improvement in understanding of mental health concerns is displayed by this dual method.

In [8], the authors used a mixed LSTM-RNN structure in an attempt to obtain optimal accuracy. Principal Component Analysis (PCA), a dimensionality redemption method was used to imagine features. While the RNN model shown two dense layers, the LSTM model included two hidden layers. This framework is distinguished between depressive and non-disdaining materials with 99% accuracy rate.

Hybrid model has also focused on research efforts; An example is [9], which added to CNN and bidarettle LSTM (bilstm). With 94.28% accuracy rate in identifying mental stages from Twitter data, this study found a remarkable language difference between depressed and non-disdain social media manifestations.

A recent study in [10] used a LSTM deep recurring network to analyze bungalow social media data. The study improved the accuracy of depression diagnosis by adjusting the hyperperameter, offering useful information to researchers and mental health professionals. Additionally, the deep-knowledge-infrastructure depression identification system was presented by [11], which laid a strong emphasis on integrating domain knowledge for better identity.

In summary, the research displays the significant impact of the LSTM model in the NLP, especially when it comes to using lesson data analysis to solve complex issues such as mental health. Researchers have continued the ability of LSTM architecture through integration of creative approaches and various data sources, which is opening the door for deep insight and better prediction skills in the field of natural language processing. Scholars' research and significant contribution to real-world applications are estimated by continuous investigation of these technologies [12–13].

3. PROPOSED METHOD

3.1 Recurrent Neural Networks (RNN)

The architecture of RNNs is designed to incorporate contextual information from previous time steps when processing a word at step t. Specifically, the RNN is composed [14] of a series of sequentially linked cells, where each cell not only processes the current word but also integrates information from the output of the preceding cell. This interconnected structure can be represented both as a recursive process within individual cells and as a linear sequence of cells, highlighting the dynamic stream of data in the network as presented in figure 1. In the realm of natural language processing, it is essential to address the variability in text length across different data instances.

To ensure uniformity in sequence processing, RNNs standardize these sequence dimensions to a predetermined fixed length. For smaller sequences compared to this fixed size, padding is applied to expand them, ensuring that they align with the required dimensions. Conversely, the sequence that exceeds the specified length undergoes tranquility to eliminate additional data, which facilitates more efficient and consistent input for RNN during training and estimates processes.

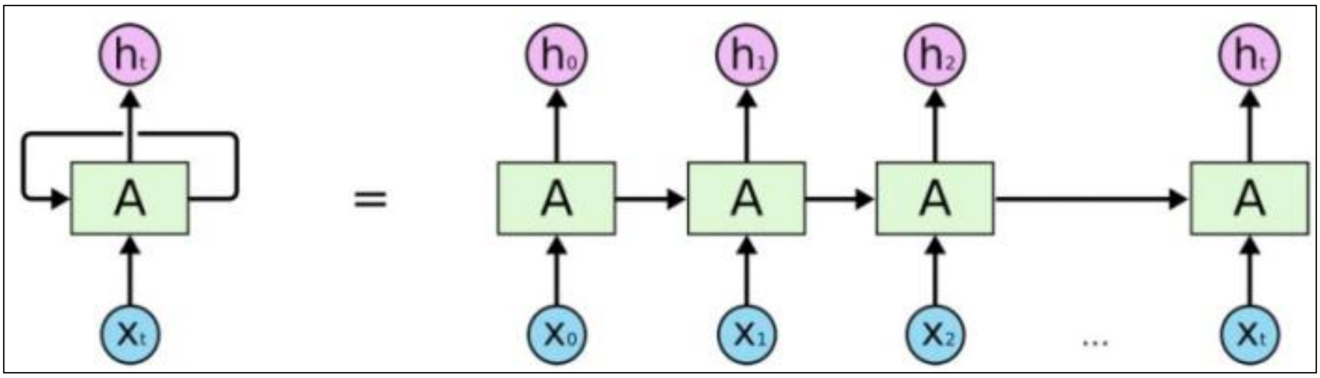


Figure 1. RNN architecture

3.2 Word Embedding

Along with machine learning models -deep learning framework demands changes in numerical formats of text data to facilitate effective processing and analysis. This requirement arises from the underlying boundaries of the algorithm that cannot explain the raw text or wire. Traditional functioning has addressed this challenge by assigning indices from predetermined vocabulary [15] in each word, in which a-hot encoding is a notable example. In this encoding scheme, each word matches a unique binary vector, in which all components are initiated up to zero, except for the specified position for a specific term, which is marked by one. Alternatively, the vector space model (VSM) conceptualize the text as a vector in a multi-dimensional location, where the dimensions correspond to the words and reflect their frequencies within the value document [16].

While these traditional techniques are effective and simple, they display critical shortcomings, such as high altitude, challenges related to computational efficiency when managing the semantic equality between the words, and rare datasets. As a result, these issues can adversely affect the performance of the classification model. To reduce these boundaries, the word embedding has emerged as a sophisticated solution, which enables meaningful word representation that surrounds both syntactic and semantic relations, which leads to the results.

3.2.1 Word2Vec

In a-hot encoding, individual words are represented as different characteristics, which lacks a relationship. Conversely, Word2VEC wants to generate vector representations that surround the intellectuals between words. This is realized through two primary methodologies: Uninterrupted Bag of Words (UBOW) and Skip-Gram. Both approaches utilize the local linguistic context, characterized by a specified window of neighboring words.

UBOW formulates word vectors by forecasting the target word from its contextual neighbors, whereas Skip-Gram operates conversely, using the target word to infer the surrounding contextual words. This duality enhances the comprehension of word semantics in natural language processing.

3.2.2 Glove

GloVe embedding utilize vector differences to express the relationship between two words through their co-occurrence probabilities. The technique uses the purpose function of a weighted minimum classes, which is depicted by J, which wants to reduce the difference between two words vectors and the co-event frequency of their dot product.

$$J = \sum_{i,j=1}^{V} \quad f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2 \qquad \text{Eq 1.}$$

Glove model waiting function [17] Each word of this structure has a vector and a prejudice word connected to it. The term reference is comparable. The depiction of the model of word associations is improved by determining the frequency of co-events between a goal word and its context and is improved using a waiting function to modify the impact of rare or highly frequent co-events.

Such a word embedding uses a vector version to encrypted overlapping probability ratio between two words. By using the purpose of a weighted at least sections, the glove reduces the difference between the number of co-events of two words and the dot products of their vector [17].

Gloves use the co-event data of a huge corpus to catch the cemented link between embedding words. This word improves understanding of models of meaning, which is crucial in interpreting the nuanced language often found in social media posts related to depression. Words with similar meanings tend to have similar GloVe embedding. This property allows the model to generalize better from seen to unseen words with similar meanings, enhancing its ability to classify depression-related content accurately even when exposed to new or less frequent words. By bridging the gap between linguistic understanding and machine learning, the proposed approach provides a strong structure to analyze social media text, contributing to the broader goal of using technology for mental health analysis and support.

3.2.3 Fasttext

The skip-gram model faces a remarkable obstacle due to the use of different vector representations for individual words, ignoring internal sub-structures. Fasttext addresses this limit effectively by increasing the skip-gram framework with character-level N-Gram. In this increased functioning, a word is decomposed into different character N-Gram, which facilitates its representation. For example, the term "extract" can be broken into components such as "Ext," "RAC," and "Act". As a result, the overall word embedding in N-Gram [18] is derived from the collected vector. In addition, Fasttext's architecture allows for the representation of out-of-vocabulary words to take advantage of existing N-Gram, which improves modeling of immoral or novel words [19–20].

3.3 Data Pre-processing

The dataset undergone a careful cleaning process, during which irrelevant components, including web addresses, special symbols and numeric entries, were abolished. Subsequently, text related messages were divided into individual tokens, in which normal stop words were excluded to focus on the most important terms. To streamline the vocabulary, lemmatization techniques were employed, thereby reducing the diversity of words and contributing to improved model performance. To remedy a minor class imbalance—characterized by approximately 53% of instances categorized as "not depressed" and 47% as "depressed"—a resampling method was implemented. In particular, to equal its representation with the majority class, the minority class ("sad") was extended by the creation of synthetic data. A-gap encoding was then used to convert processed text data into a numerical format. For the convenience of model training, evaluation and adaptation, the dataset was eventually divided into the best for training, testing and verification with a distribution ratio of 60:20:20.

3.4 Model LSTM and RNN

Recurrent neural networks (RNNs) effectively manage the retention of information across sequences of data. This capability is facilitated by a unique memory cell that enables the preservation of critical information over significant time intervals. Central to RNN architecture is the cell state, which operates on a selective basis to either maintain or eliminate data depending on the recent response and the unseen state derived from earlier time steps. The structural design of a traditional RNN cell, alongside its single-cell configuration.

3.4.1 Model Training

With a batch size of 32, each RNN variation was trained over 10 ages on training dataset. The model parameters were updated using adam optimizer, an adaptive learning rate adaptation algorithm. Damage functions used to determine the discrepancy between the real and approximate label were binary cross-entropy.

3.4.2 Model Evaluation

The test dataset was used to evaluate the performance of each model. The F1-score, accurate, recall and accuracy were among the assessment matrix. Confusion matrices of confusion were made for each model to examine the findings. The best model to divide the social media text into depression and non-disdaining categories was found with the use of in visualization, which also helped compare the performance of RNN versions.

4. RESULTS AND DISCUSSION

Research used the python programming language as a combination with clag platforms, which employs the T4 GPU for cloud computing abilities. This functioning systematically compares various RNN architecture applied to datasets obtained from social media text. Results achieve significant insights into comparative efficacy of uneven RNN models in terms of detection of depression through textual analysis.

4.1 Dataset

Reddit is known as a huge collection of public posting from users, including various types of mental health problems including depression, known as self-reported mental health diagnosis (SMHD) dataset. This dataset contributes to people matched without diagnosis, providing intensive understanding of interaction around mental health from January 2006 to December 2017. Dataset in mental health research has been exposed to its remarkable macro-average accuracy 95.8% and a minimum accuracy of 90% for anxiety. Classification [21]. Using SMHD dataset, we created depression-specific datasets for our investigation and matched them with control groups. This approach features training and evaluation of Sbert-CNN models, aimed at identifying Reddit users who demonstrate signs of depression. The original structure of the SMHD dataset allowed us to maintain a clear division into training, validation, and testing subsets. This segregation is crucial for ensuring that model training, parameter adjustments, and evaluations are conducted on distinct data sets, thereby enhancing the model's predictive performance. This structured approach not only enriches our understanding of depression discourse on social media however similarly provides a robust basis for developing effective identification models.

4.2 Training Process

20% of Reddit dataset is used for testing, and the remaining 80% is used for training. RNN test training does not use dataset, and vice versa. Ten percent of training data is used for data verification. The volume of each dataset is divided randomly using an automated data division.

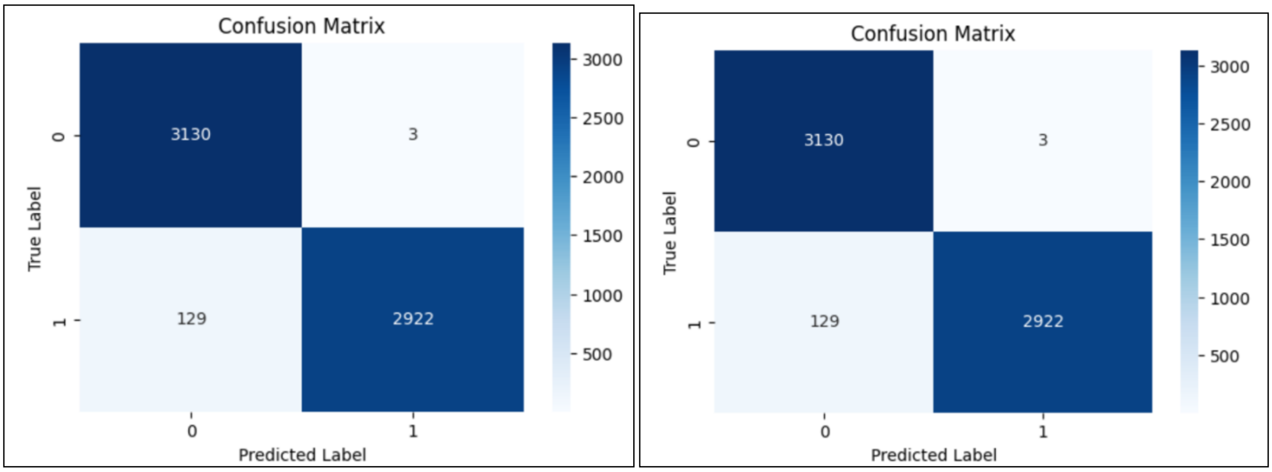
4.3 RNN Glove model

Matrix obtained from pre-processing on hyper-parameter embedding matrix input, Relu and Thani's activism at Hidden Gate, Softmax Activation at Output Gate, Optimizer Adam and RMSPROP, Dropout 0.5 and APOCH 50, with 0.5 and Apoc Is done. The classification training process embeding feature global vector (glove) using 300 dimensions. Each of the eight models is trained with tuning rates of 0.001 and 0.0001. The rate or speed of the learning model is controlled by the hyperpimeter learning rate. In particular, it controls how many divisions their model weight is updated in mistakes, each time they are updated, for example, each training example at the conclusion of the example batch. Perhaps the most important hyper-perimeter is the learning rate.

4.4 Results

Using the python on the Kaggle platform, the performance of the model suggested on the Reddit dataset was evaluated. The results obtained show how effective the RNN model enhanced with glove embeding. With 97.84% training accuracy and 97.22% test accuracy, the model demonstrated very low overfiting and high levels of stability throughout the training and testing stages.

The purpose of the introduction is to explain to readers all details of this research. Its role is to arouse the reader's attention and let the reader have a general understanding of the paper. The contents to be described in the introduction are roughly as follows: (1) The rationale, purpose and background of the research. Including the question, the research object and its basic characteristics, what work has been done by the predecessors on this issue, what are the deficiencies; what problems are expected to be solved, what is the role and significance of the solution; what is the background of the research work. If you want to answer a lot of questions, you can only take a brief explanation. Usually, you can explain one problem in one or two sentences. (2) Theoretical basis, experimental basis and research methods. If you follow the known theory, principles, and methods, just mention a paragraph, or note the relevant literature. If a new concept or term is to be introduced, it should be defined or clarified. (3) The expected results and their status, role and significance should be written in a natural, general, concise and precise manner. In the introduction, diagrams, tables, and formulas are generally not allowed.



(a) Confusion Matrix for Training ***(b) Confusion Matrix for Testing***

Figure 2. Confusion Matrix for training and testing classification

As seen in Figure 2, a completely classification report further displays the flexibility of the model. Accurate, recall and F1-score for class 0, which represents unrelated positions from depression, were 0.95, 0.99 and 0.97 respectively. For class 1, accurate, recall and F1-score, which represented posts related to depression, were 0.99, 0.95 and 0.97 respectively. With a weighted average F1-score of 0.97 and a macro average, the overall accuracy of the model was 97%.

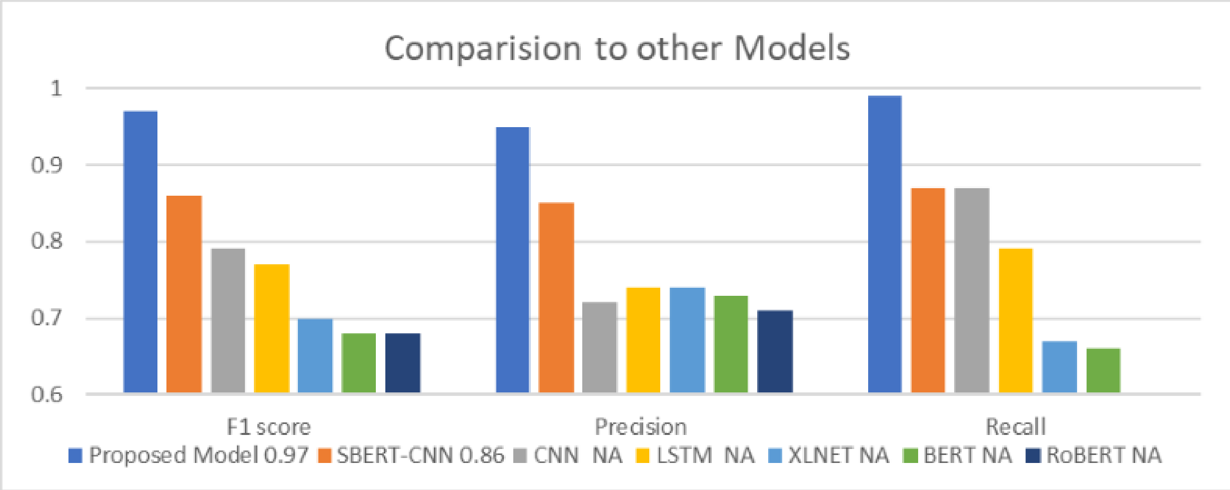


Figure 3. Results comparison of proposed model.

Table 1 highlighted these findings in confusion matrix for training and testing data. This results shows how well the model can distinguish between posting with some errors that are related to depression and which are not. While high recall for class 0 guarantees some false negatives in unrelated predictions, high precision for class 1 guarantees that the model successfully detects material related to depression. For applications in mental health, where both excessive and insufficient detection can have serious consequences, this balance is necessary. Comparing models suggested with other models that are currently in use on a general Reddit dataset, are displayed in Figure 3 and Table 2. The results obtained show how effective the RNN model enhanced with glove embedding.

Table 1. Classification report

	precision	recall	f1-score	support
Class 0	0.95	0.99	0.97	767
Class 1	0.99	0.95	0.97	780
accuracy			0.97	1547
macro avg	0.97	0.97	0.97	1547
weighted avg	0.97	0.97	0.97	1547

Table 2. Performance of Proposed model compared to other Existing Models

Model	Accuracy	F1 score	Precision	Recall
Proposed Model	**0.97**	**0.97**	**0.95**	**0.99**
SBERT-CNN [22]	0.86	0.86	0.85	0.87
CNN [23]	NA	0.79	0.72	0.87
LSTM [23]	NA	0.77	0.74	0.79
XLNET a [24]	NA	0.70	0.74	0.67
BERT a [24]	NA	0.68	0.73	0.66
RoBERT a [24]	NA	0.68	0.71	0.60

5. CONCLUSION

Using the Natural Language Processing (NLP) tool, the task examines the creation of an accurate model for early detection of depression. The lesson data of social media channels was used in comparative studies of five different long-term short-term memory (LSTM) network. This study shows how gloves embedding have a major effect how well the recurrent nerve network (RNN) classify depression. The suggested structure machine provides an intensive strategy to check social media information by fusing linguistic insight with learning techniques, supporting the comprehensive effort to use technology for mental health evaluation and assistance. After intensive evaluation, the RNN model with gloves embedding showed better performance than the baseline model without pre-educated embedding. The results throw light on how important it is to use advanced words such as gloves to improve the classification accuracy of texts about depression, showing that they are accompanied by initial identity and early treatment of mental health issues.

REFERENCES

[1] Naslund, J. A., Tugnawat, D., Anand, A., Cooper, Z., Dimidjian, S., Fairburn, C. G., ... & Patel, V. (2021). Digital training for non-specialist health workers to deliver a brief psychological treatment for depression in India: Protocol for a three-arm randomized controlled trial. Contemporary clinical trials, 102, 106267.

[2] Keerthan Kumar, T. G., Himanshu Dhakate, and Shashidhar G. Koolagudi. "IIMH: Intention Identification In Multimodal Human Utterances." In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, pp. 337-344. 2023.

[3] Vydiswaran, V. V., & Reddy, M. (2019). Identifying peer experts in online health forums. BMC medical informatics and decision making, 19, 41-49.

[4] Gupta, S., Goel, L., Singh, A., Prasad, A., & Ullah, M. A. (2022). Psychological analysis for depression detection from social networking sites. Computational Intelligence and Neuroscience, 2022.

[5] Singh, D., & Wang, A. (2016). Detecting depression through tweets. Standford University CA, 9430, 1-9.

[6] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. IEEE Access, 7, 44883- 44893.

[7] Solieman, H., & Pustozerov, E. A. (2021, January). The detection of depression using multimodal models based on text and voice quality features. In 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus) (pp. 1843-1848). IEEE.

[8] Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep learning for depression detection from textual data. Electronics, 11(5), 676.

[9] Kour, H., & Gupta, M. K. (2022). An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bidirectional LSTM. Multimedia Tools and Applications, 81(17), 23649-23685.

[10] Uddin, A. H., Bapery, D., & Arif, A. S. M. (2019). Depression Analysis from Social Media Data in Bangla Language Applying Deep Recurrent Neural Networks.

[11] Zhang, W., Xie, J., Liu, X., & Zhang, Z. (2023). Depression Detection Using Digital Traces on Social Media: A Knowledge-aware Deep Learning Approach. arXiv e-prints, arXiv-2303.

[12] Sushma, S. A., and Keerthan Kumar TG. "Comparative Study of Naive Bayes, Gaussian Naive Bayes Classifier and Decision Tree Algorithms for Prediction of Heart Diseases." (2021).

[13] Varun, J., E. S. Vishnu Tejas, and T. G. Keerthan Kumar. "Performance Analysis of Machine Learning Algorithms Over a Network Traffic." In International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020, pp. 1-10. Springer Singapore, 2021.

[14] Wei, Daqian, et al. "Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report." Energies 10.3 (2017): 406.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc., pp. 1–12, 2013.

[16] M. Azam, T. Ahmed, F. Sabah, F. and M.I. Hussain, "Feature Extraction based Text Classification using K-Nearest Neighbor Algorithm". IJCSNS Int. J. Comput. Sci. Netw. Secur, 18, pp. 95-101, 2018.

[17] HANSON ER, "Musicassette Interchangeability. the Facts Behind the Facts," AES J. Audio Eng. Soc., vol. 19, no. 5, pp. 417–425, 1971.

[18] P. Association for Computational Linguistics, E. Grave, A. Joulin, and T. Mikolov, "Transactions of the Association for Computational Linguistics.," Trans. Assoc. Comput. Linguist., vol. 5, pp. 135–146, 2017.

[19] C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," ACM Int. Conf. Proceeding Ser., pp. 37–46, 2020, doi: 10.1145/3443279.3443304.

[20] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, "Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews," Math. Probl. Eng., vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/5536560.

[21] Cohan, A., et al., SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions, 2018.

[22] Chen, Z., Yang, R., Fu, S., Zong, N., Liu, H., & Huang, M. (2023, June). Detecting Reddit Users with Depression Using a Hybrid Neural Network SBERT-CNN. In 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI) (pp. 193-199). IEEE.

[23] de Souza, V.B., J.C. Nobre, and K. Becker, DAC Stacking: A Deep Learning Ensemble to Classify Anxiety, Depression, and Their Comorbidity From Reddit Texts. IEEE Journal of Biomedical and Health Informatics, 2022. 26(7): p. 3303-3311.

[24] Dinu, A. and A.-C. Moldovan. Automatic detection and classification of mental illnesses from general social media texts. in Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021). 2021.