Detecting early signs of depression in twitter posts through LSTM-based sentiment analysis

Rupesh Mahajan[1], *Nikita* Bangar[1], *Aditi* Vaibhav Apte[1*], *Nupur* Mukund Chaudhari[1], *and Jayshree* Santosh Parde[1]

[1]Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India

Abstract. Depression is one of the most common mental health disorders, yet a large proportion of the cases are left undiagnosed due to associated social stigma, limited awareness, and lack of timely clinical intervention. The unprecedented growth of social media has encouraged users to frequently express their emotions on the social Web, thus opening an opportunity for the early detection of psychological distress from their digital footprints. Most of the available research works target single-language sentiment analysis or few classifier benchmarking, and thus there is a lack in multilingual and comparison-based deep learning-based studies. This work attempts to fill such a research gap by proposing a supervised machine learning framework for the identification of depression indicators from both English and Marathi tweets. The proposed approach integrates the standard NLP preprocessing—tokenization, stop-word removal, stemming, and TF-IDF and Word2Vec-based feature extraction—with training multiple classifiers, namely Logistic Regression, SVM, Random Forest, and LSTM. The experimental results demonstrate that the LSTM model yields state-of-the-art performance compared to other traditional models due to its natural ability to capture contextual dependencies of texts. The findings have important implications for the social media-driven method for early mental health screening and public health monitoring and digital psychological intervention systems.

1 Introduction

Mental illnesses like depression, anxiety, and stress are wide-spread in the present times. Depression, as estimated by the World Health Organization (WHO), has more than 280 million sufferers in the world and is a main determinant of disability as well as suicide. The conventional method of diagnosis employs clinical interviews and psychological tests, which take time and are usually not available.

Social media sites like Facebook, Twitter, and Reddit have been a goldmine source of information in the past few years to understand the activity, temper, and emotions of users. Users share their opinions and feelings without any restriction and even share clues of mental

* Corresponding author: aditiapte1504@gmail.com

illness sometimes. It is here that machine learning and NLP can be utilized to identify depression using text data.

The paper is centered on creating a multilingual depression detection system between Marathi and English tweets. The aim is to create a system capable of analyzing textual indicators, emotional tone, and linguistic styles in trying to identify whether or not a user is predisposed to being depressed. The study is also centered on the possibilities of merging multilingual data for enhanced contextual and cultural sensitivity. [1,2,3,4,5]

2 Literature Review

Other research also considered the use of NLP and deep learning approaches to depression detection from available text data on social sites.
Orabi et al. have experimented with deep learning algorithms on Twitter data for identification of depressed people, and their result also indicates that the deep learning-based solution outperforms the existing classifiers in understanding context.[3]
In their study, Coppersmith et al. (2018) engaged in the linguistics and verbatim emotional cues with regards to social media messages as a screening tool for suicidality. A review by Coppersmith and Néron de Pontbriand discusses the use of NLP in assessing mental well-being.[5]
Meanwhile, Singh et al. (2022) have proposed an embedded LSTM framework to detect depression, where learning in a sequence improves the accuracy of text-based feature learning for this classification task. [1]
Sridharan et al. applied deep learning models for depression detection in the users of social platforms based on polarity and expressions of emotions in user-generated content.[2]
Kour & Gupta 2023 proposed a hybrid CNN & BiLSTM, which outperformed the conventional deep learning models for depression prediction.[7]
However, most of these works focus on English datasets, and hence the applicability of depression detection models is restricted in multilingual communities. The solution proposed in this work tackles the problem by giving reasonable importance to multilingual analysis.

3 Methodology

3.1. System Overview

System design involves the following key steps:
1.Data collection from Twitter.
2.Preprocessing and cleaning of English-Marathi text.
3.TF-IDF and word embedding feature extraction.
4.Learning the model by supervised models.
5.Testing with standard metrics.
[17,18]

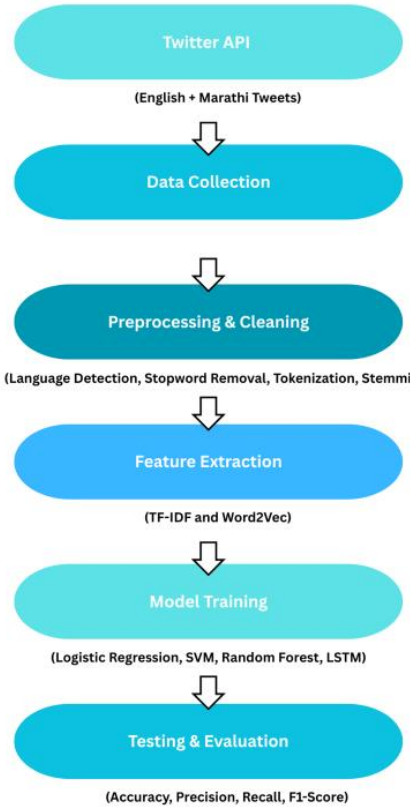

Fig 1. System Architecture

Figure 1 shows the complete system architecture used in this study, illustrating the flow from data collection to preprocessing, feature extraction, model training, and final evaluation.

3.2. Data Collection

We have collected tweets from the Twitter API using a list of depression-related hashtags and keywords like #depression, #sad, #lonely, #उदास, #मन खट्टू आहे, and other such words and phrases through which one conveys a state of being sad or some other sort of negative emotions. In total, our dataset contains approximately 10,000 tweets gathered in three months. Tweets were filtered to include both English and Marathi languages in order to capture a wider cultural and linguistic context.

Duplicate and retweeted posts were removed to assure the quality of the data, keeping only original content. Afterwards, the corpus was manually annotated into two major categories:

Depressed - 1: Those tweets that contain explicit or implicit signs of depressive or negative emotional states.

Not Depressed (0): The tweets are either neutral, positive, or generally unrelated to depression.

Each of the entries in the tweet consists of the text of the tweet, language label that is either English or Marathi, sentiment polarity, and timestamp. To minimize the bias, annotations were conducted manually by three human reviewers, who also demonstrated strong inter-annotator agreements confirmed by Cohen's Kappa at 0.87. The dataset consisted of approximately 60% English tweets and 40% Marathi tweets.

Additional metadata included the number of likes and retweets, as well as posting frequency, in order to support further behavioral and temporal analysis. This dataset is

bilingual in nature and forms the basis of the training of the model; it represents a realistic discourse on social media with respect to mental health in the Indian context.

3.3. Data Preprocessing

Text preprocessing is required to cleanse unstructured noisy data. The following has been used:

1.Language Detection: Detecting English or Marathi text.
2.Lowercasing: Converting everything to lower case for consistency.
3.Stop-word Removal: Removing usual words (e.g., "the," "आहे," "मला") without any meaning.
4.Tokenization: Sentence splitting into tokens.
5.Stemming/Lemmatization: Converting words to the root.
6.Noise Removal: Removing the hashtags, URLs, mentions, emojis, and special characters.

This helps in both the English and Marathi text to obtain normalized so that there is proper feature representation.

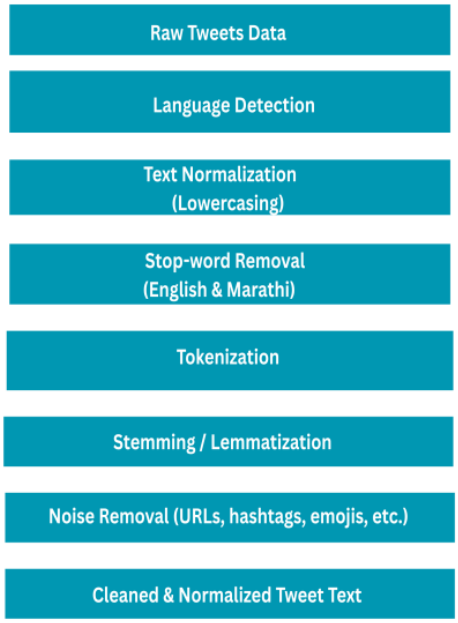

Fig 2. Data Cleaning and Preprocessing Pipeline

Figure 2 presents the preprocessing and cleaning pipeline applied to the English–Marathi tweets, highlighting the sequence of steps from raw text to normalized data. [10,18]

3.4. Feature Extraction

For text conversion into numerical representation for machine learning models, two methods of feature extraction were employed:

1.TF-IDF (Term Frequency–Inverse Document Frequency): Writes down word importance with regards to document frequency.
2.Word2Vec Embeddings: Maps semantic meaning of words to a vector space in real numbers, enhancing contextual meaning. [18]

3.5. Model Development

Four supervised learning models were trained and compared:

1.Logistic Regression (LR): Basic baseline classifier for binary prediction.

2.Support Vector Machine (SVM): Well-suited for high-dimensional text classification.
3.Random Forest (RF): Ensemble classifier consisting of many decision trees.
4.LSTM (Long Short-Term Memory): Deep learning model that can understand word sequences and emotional context.

Hyperparameters were optimized by grid search to get optimal accuracy.
[1,7,18]

Fig 3. LSTM Model Architecture

Figure 3 illustrates the architecture of the LSTM model used in this study, including the embedding layer, LSTM cells, dense layer, and final sigmoid classifier for binary prediction.

3.6. Evaluation Metrics

Performance was measured using:
1.Accuracy
2.Precision
3.Recall
4.F1-Score
5.Confusion Matrix

Cross-validation was performed to ensure the generalizability of results. [8,18]

4 Result Analysis

After training and testing with the bilingual data set, we found that deep learning models outperformed the other models in performance.

The LSTM model exceeded the other models when characterizing the language and emotional differences between English and Marathi tweets.

Some of the primary findings of this project were:

The Marathi tweets greatly assisted in advancing sentiment understanding, supporting the idea that emotion classification was enhanced by the regional context.

The system was able to differentiate between neutral sadness and the differences in language patterns on clinical depressive language.

Table 1. Performance Comparison of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	84%	83%	80%	81%
SVM	87%	86%	85%	85%
Random Forest	88%	87%	86%	86%
LSTM	91%	90%	92%	91%

The results in Table 1 show that traditional machine learning models, including Logistic Regression, SVM, and Random Forest, work relatively well for the depression detection tasks. However, the LSTM model yields the best metrics: accuracy at 91%, precision at 90%, recall at 92%, and an F1-score of 91%. This improvement is attributed to the LSTM's ability to capture long-term contextual dependencies within text sequences, making it more effective for detecting subtle emotional cues in depressive language.

5 Discussion

The proposed approach presents how supervised machine learning algorithms work successfully for identifying depression through the analysis of written communications. The automated computer system implements multi-bilingual NLP strategies to formulate the correct output for each user category. The organization is currently encountering numerous impediments to address prior to moving forward.

The partitioning of messages into tokens is difficult to filter for Tweets switching between English and Marathi language for short periods within the same Tweet message.

Also, the performance of the classification system performs poorly because the minority class did not have a sufficient labeled training sample.

Essentially the difficulty lies with the data analysis process lacking enough contextual cues for detail.

Overall, the system demonstrates great promise in detecting mental health conditions at their first indications. The system also tracks mental health through the use digital well-being tools, and healthcare systems and analytics platforms.

6 Future Scope

The envisioned future directions of the research may consist of the following items:

1. larger and more diverse datasets from other social media platforms and languages".
2. employing transformer models for better context understanding, such as BERT, mBERT, or RoBERTa.
3. monitoring tweets in real time for the purpose of providing an early warning to intervene.
4. integrating the technology into mobile applications or chatbots as an avenue for providing support for mental health issues.

5. partnering with psychologists to ethically deploy the system and conduct clinical validation.

The advancements discussed above represent the potential for the development of smart systems to assist mental health practitioners, social welfare organizations and counselors in their efforts to identify at risk individuals.

7 Conclusion

This research developed a bilingual supervised machine learning framework that could detect cues related to depression from both English and Marathi tweets. The proposed system used standard NLP preprocessing along with TF-IDF and Word2Vec for feature extraction, followed by multiple classifier training involving Logistic Regression, SVM, Random Forest, and LSTM. Among these, through an extensive experiment, LSTM performed the best since it can retain contextual dependencies of high order in textual data. This empirical result shows that deep learning models are more useful in sentiment-based mental-health prediction tasks. The use of data in multiple languages enhanced diversity and usability in regions with linguistic diversity. Therefore, this framework has immense potential in developing early intervention tools for the support of mental-health monitoring and awareness and timely digital psychological intervention.

References

1. Singh, J., Wazid, M., Singh, D. P., & Pundir, S. (2022). *An embedded LSTM-based scheme for depression detection and analysis*. Procedia Computer Science, 215, 166–175.
2. Sridharan, S., Aishwarya, S., Aishwarya, S. P., & Santhiya, S. (2023). *Depression detection in social media users using deep learning*. American Journal of Psychiatric Rehabilitation, 26(1), 45–58.
3. Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D. (2018). *Deep learning for depression detection of Twitter users*. In CLPsych Workshop (pp. 88–97).
4. Devi, T. J., & Gopi, A. (2024). *The evaluation of deep learning models for detecting mental disorders*. International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 12(2), 345–352.
5. Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). *Natural language processing of social media as screening for suicide risk*. Biomedical Informatics Insights, 10, 1–11.
6. Narayan, U., & Kumar, D. (2024). *Sentiment analysis using transformer-based model*. International Journal of Innovative Research in Computer Science & Technology (IJIRCST), 12(S1), 210–218.
7. Kour, N., & Gupta, D. (2023). *A hybrid deep learning approach for depression prediction using CNN–BiLSTM*. Multimedia Tools and Applications, 82, 40963–40988.
8. Bokolo, A. J., & Liu, Y. (2023). *Deep learning-based depression detection: Machine learning vs transformers*. Electronics, 12(21), 4396.
9. Bendebane, H., et al. (2023). *Multi-class deep learning for depressive/anxiety disorders*. Algorithms, 16(12), 543.
10. Sajib, M. H., Hasan, M. R., & Rahman, M. M. (2024). *Depression detection from Bangla text using RNNs*. arXiv:2412.05861.
11. Singh, R., Kumar, A., & Sharma, V. (2024). *Sentence-BERT ensemble for depression detection*. arXiv:2409.13713.

12. Ta, N., Alhassan, A., & Burley, D. (2025). *Detecting signs of depression*. Safety and Health at Work, 16, 100827.
13. Rahman, M. M., Sultana, T., & Akter, S. (2024). *Depression detection through explainable AI*. arXiv:2404.13104.
14. Abirami, S., Priyadharshini, R., Rajeswari, S., & Kumar, P. (2025). *Explainable AI-driven depression detection*. Frontiers in Artificial Intelligence, 5, 1627078.
15. Adewumi, A. O., Owoeye, O., & Misra, S. (2021). *A hybrid approach to detecting depression*. arXiv:2106.10485.
16. Nadeem, M., Iqbal, A., Malik, H., & Rehman, S. (2025). *Early detection of depression*. SES Journal, 6(1), 1–12.
17. Goel, N., Anand, R., Amoli, A., Chuki, K., Sharma, P., & Harnal, S. (2024). *Automated depression detection system*. International Journal of Engineering Research & Technology (IJERT), 13(05), 245–252.
18. Ahmed, S., Rakin, S., Ibn Waliur, M. W., Islam, N. B., Hossain, B., & Akbar, M. M. (2024). *Depression detection from Bangla using RNNs*. arXiv:2412.05861.
19. De Choudhury, M., Counts, S., & Horvitz, E. (2013). *Predicting depression via social media*. In ICWSM 2013 (pp. 128–137).
20. Rahman, M., Chowdhury, S., Khan, T., Hossain, M., & Das, A. (2025). *Depression and suicidal tendency detection*. Behavioral Sciences, 15(3), 352.