EasyChair Preprint

№ 12731

Depression Detection Using Machine Learning

Sai Kotesh Deekonda, Abhinash Atchala, Vishal Bommerla and Rajesh Kumar Raj Purohit Derani

March 27, 2024

DEPRESSION DETECTION USING MACHINE LEARNING

Sai Kotesh Deekonda
Department of Computer Science Engineering and Technology
Parul Institute of Engineering and Technology
Vadodara, India
200303124193@paruluniversity.ac.in

Abhinash Atchala
Department of Computer Science Engineering and Technology
Parul Institute of Engineering and Technology
Vadodara, India
200303124121@paruluniversity.ac.in

Vishal Bommerla
Department of Computer Science Engineering and Technology
Parul Institute of Engineering and Technology
Vadodara, India
200303124155@paruluniversity.ac.in

Rajesh kumar raj purohit Derani
Department of Computer Science Engineering and Technology
Parul Institute of Engineering and Technology
Vadodara, India
200303124209@paruluniversity.ac.in

***Abstract*— Millions of individuals all around the world suffer from depression, a significant mental health problem. The effectiveness of treating depression can be greatly enhanced by early identification and intervention. In this paper, we create a robust and effective model for depression identification by combining the capabilities of deep learning with conventional machine learning methods. We get information from a range of sources, including blogs, internet forums, and social networking sites. We use natural language processing methods to preprocess the input and extract pertinent information. The underlying patterns in the data are then discovered using a convolutional neural network (CNN) and a recurrent neural network (RNN). It is now possible to realise an automated system that can assist in identifying depression in individuals across a range of ages. Researchers have been searching for methods to accurately diagnose depression in order to detect it. Several investigations have been suggested in this context. In this work, we review a number of prior investigations that employed machine learning (ML) and artificial intelligence (AI) to identify depression. We integrate the output of these models with well-known machine learning techniques, such as decision trees and support vector machines (SVM), to increase the precision of depression diagnosis. Our findings demonstrate that in terms of accuracy, precision, and recall, the suggested hybrid model performs better than the methods currently in use. By the early and precise identification of depression, prompt intervention, and improved patient outcomes, this paradigm has the potential to completely transform the area of mental health.**

Keywords—Depression screening, Mental health assessment, Mood analysis Emotional recognition, Behavioral patterns, Predictive analytics.

I. INTRODUCTION

An excessive number of individuals globally suffer from depression, a mental health condition. It can result in symptoms that have a major negative influence on a person's quality of life, such as persistent depressive moods, lack of interest in activities, changes in food, and irregular sleep patterns. Identifying depression early is essential for effective treatment and improved outcomes. Machine learning has emerged as a promising approach for depression detection and screening. Large datasets may be analysed by machine learning algorithms, which can also spot patterns that are hard for people to see. In the context of depression, machine learning algorithms can analyze various data sources, such as patient self-reports, electronic health records, and social media data, to detect patterns and identify potential cases of depression. The development of depression detection using machine learning involves training algorithms to identify features associated with depression, such as changes in mood, behavior, and speech patterns. These features are then used to classify patients as either having depression or not. To train these algorithms, large amounts of data are needed, including patient data and clinical records. Overall, depression detection using machine learning has the potential to improve early diagnosis, treatment, and outcomes for individuals with depression. By leveraging the power of machine learning, we can help reduce the burden of this debilitating condition on individuals and society as a whole.

II. LITERATURE SURVEY

This paper explores methods for detecting depression using machine learning (ML) and deep learning (DL).The KBRS system uses text analysis to identify negative sentences with high accuracy (0.89-0.90).Social media analysis can also be used to detect depression with some success (0.70-0.85 accuracy).Different techniques like CNNs, LSTMs, and SVM classifiers are used in various studies. Text analysis often focuses on identifying keywords related to depression. Some studies incorporate audio features in addition to text data. The goal is to develop methods that can assess depression severity[1]. This study explores using machine learning to diagnose depression. The authors created a system called AiME that uses deep learning to analyze speech, facial expressions, and self-reported information to identify depression with high accuracy (87.3%). This technology could be a valuable tool for mental health professionals by providing a more objective assessment and allowing for earlier intervention[2].This article proposes a system to identify depression in social media users by analyzing their posts. The system uses machine learning, specifically Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) classifiers, to categorize tweets as depressive or not depressive. The system is trained on a dataset of emotion-labeled text and then analyzes social media profiles. If more than 75% of a user's tweets are classified as depressive, the system flags the profile for potential depression. The authors reference a study by Muzafar et al. that found similar results using machine learning to detect depression[3]. This study investigates using machine learning to detect depression in children aged 4-17. They analyze data from a mental health survey and identify key features of depression in this age group. The Random Forest machine learning model

achieved the best results in accuracy and speed at identifying depression. This approach has the potential to improve early detection of depression in children[4]. This article discusses using deep learning (DL) to detect depression. DL can be used to analyze text and video data to identify signs of depression. CNNs are a type of deep learning model that can be effective for this task. Early detection of depression is important to allow for early treatment. There are challenges in using DL for depression detection, such as the difficulty of defining depression stages and the lack of large public datasets. The article also proposes a system that uses facial expressions to detect depression and suggests interventions for people who are detected as depressed[5]. This study explored using mobile devices with passive sensing to monitor mental health in Nepal. Teenagers, young mothers, and their families with mental health issues participated for 6 weeks. The devices collected data on behavior, location, and social interactions. The findings showed that both participants and their families accepted the technology. The data provided insights into mental health and could be used to develop better treatments. Including families also helped reduce stigma around mental health[6]. This study explores using deep learning to detect depression from audio recordings. They built a deep learning model using audio features from people diagnosed with depression. The model achieved good results in identifying depression but needs further research. Deep learning shows promise for detecting depression using audio data[7]. This research explores using social media to identify and categorize depression severity. Here's a summarized version: **Problem:** Early detection of depression is challenging due to limitations in traditional methods. **Solution:** Analyze user-generated content (posts) and associated mood tags to detect depression and its severity. **Method:** Data is collected from online communities. Sentiment analysis (LIWC) is used to understand the emotional tone of posts within a community. Machine learning and statistics identify key characteristics that differentiate depressive content from non-depressive content. The severity of depression in a post is estimated based on the mood tag's "valence" value (more negative = more severe).**Results:** The system achieved high accuracy (over 90%) in classifying depressive posts, depressive communities, and depression severity. **Conclusion:** Social media analysis shows promise in identifying depression and its severity based on user-generated content.This approach could be valuable for mental health professionals in: Early detection of depression risk. Intervention strategies for individuals struggling with depression[8].

This study explores how to analyze emotions expressed in Azerbaijani tweets. Researchers collected 10,000 tweets and cleaned them by removing unnecessary elements like website links and hashtags. They then converted the tweets into numerical data and assigned positive, negative, or neutral labels to each tweet. Finally, they compared different machine learning algorithms to see which ones were most accurate at identifying the sentiment of a tweet. The results showed that a technique called SVM was the best at this task, and that there might be even better ways to handle the text data in the future. This research provides a valuable roadmap for future work on understanding emotions in Azerbaijani social media posts[9]. Traditional methods for detecting mood problems rely on surveys, which can be burdensome for participants. This study explores using machine learning and wearable sensors to predict mood instability. Researchers collected data from phone sensors, social media, and daily mood ratings from 30 participants over two weeks. After processing the data, they achieved promising results in predicting mood swings using a machine learning model. This suggests that passively collected data could be a valuable tool for monitoring mental health[10]. Depression is a serious mental health disorder that can impact anyone. Traditional methods for detecting depression can be difficult due to factors like self-denial and stigma. This study explores using machine learning algorithms to analyze social media posts for signs of depression. The researchers reviewed existing studies to find which algorithms performed best. They found Logistic Regression achieved the highest accuracy (nearly 99.8%) at detecting depression in social media text. This suggests that machine learning could be a valuable tool for identifying depression, potentially as a first-line screening or to support psychologist diagnoses[11].

Depression is a serious illness, not simply sadness. While some misunderstand it as a minor issue, depression requires proper treatment and support for recovery. Predicting its course is crucial for treatment decisions. This study explored using machine learning to analyze patient data and predict how long depression might last. Researchers compared five algorithms and found that Support Vector Machines (SVM) achieved the best results (85% accuracy). This suggests that machine learning, particularly SVM, could be a valuable tool for improving depression treatment plans[12]. This article explores using machine learning to understand depression. Depression is a serious health problem that significantly impacts quality of life. Traditionally, identifying factors that cause depression has been challenging. Researchers created a new method that combines different machine learning techniques to analyze complex data. This data explores the connection between quality of life and depression. The method achieved high accuracy (91%) in identifying factors that might cause depression. This approach has the potential to improve healthcare by: Helping doctors identify the root causes of depression in patients. Leading to new treatment options and experiments. Providing valuable insights for developing better mental health support systems[13]. Detecting depression early is critical, but traditional methods can be limited. This study investigates using social media analysis to identify signs of depression. Researchers created a system that analyzes social media posts and utilizes natural language processing (NLP) to understand the meaning behind the words. By combining NLP with machine learning algorithms, the system can classify posts as potentially reflecting depression. This method achieved an impressive accuracy of 96.35%, suggesting it could be a valuable tool for mental health professionals. Early detection of depression through social media analysis could help identify individuals who might be struggling and get them the support they need[14]. Speech analysis shows promise for detecting depression objectively. This study explores using machine learning to analyze speech patterns

and predict depression severity. Researchers collected speech data while participants read vocabulary words and described images. Several machine learning models were tested, with Random Forest achieving the best results (mean F1 score of 0.71). This suggests that speech analysis could be a valuable tool for diagnosing depression. Future work aims to improve the model's accuracy and develop an application to support people with depression[15]. This study explores using machine learning to analyze text responses and identify people at risk of depression. Early-stage depression can manifest through symptoms like loneliness and disinterest. Researchers compared different machine learning algorithms and found that Random Forest performed best at detecting depression risk from written text. This suggests that machine learning could be a valuable tool for early detection of depression[16].

This study proposes a new method to automatically detect depression using audio and text data. Here's the key idea: Traditional methods rely on interviews and questionnaires, which can be subjective and time-consuming. This approach analyzes speech and text patterns to detect depression more objectively. The system uses deep learning techniques to identify relevant features in both audio and text data. The researchers believe this method offers advantages over previous approaches[17]. This study investigates using social media to detect depression. Social media has become a popular way for people to share their thoughts and feelings. Researchers can analyze this data to identify signs of depression. This study collected tweets and used a machine learning algorithm to classify them as depressive or non-depressive. Two algorithms, Naive Bayes and a hybrid model, were tested. Both models achieved similar accuracy in detecting depression. This suggests that social media analysis could be a valuable tool for identifying people who might be struggling with depression[18]. This study explores using speech analysis to predict depression severity. Here's the key takeaway: Depression is a major illness affecting millions worldwide. Early detection is crucial for effective treatment. Traditional methods rely on questionnaires and interviews, which can be limited. This study investigates using machine learning to analyze speech patterns and predict depression severity from audio recordings. The researchers achieved promising results, suggesting that speech analysis has potential as a tool for depression screening[19]. This study proposes a new method to assess depression severity using multiple data sources, including text, audio, and video. Traditional methods for depression assessment often rely on a single data source, which can be limiting. This approach combines information from different modalities (text, audio, video) to get a more comprehensive picture of a person's depression level. The researchers use a special attention mechanism to focus on the most relevant parts of each data source and improve the accuracy of their assessments. They also use a technique called Self-Attention Generative Adversarial Networks (SAGAN) to generate more training data, which is important because large amounts of data are often needed to train machine learning models effectively. Overall, this study suggests that a multimodal approach using deep learning has the potential to improve the accuracy of depression assessment[20].

III. METHODOLOGY

1.Data collection: Collect a large amount of data related to depression from various sources such as questionnaires, surveys, social media, and clinical records.
2. Data pre-processing: Clean the collected data by removing any missing values, duplicates, and outliers. Transform the data into a format suitable for analysis.
3. Feature selection: Select the most relevant features (e.g., symptoms, demographic information) that can help identify depression from the pre-processed data.
4. Data splitting: Split the pre-processed data into training and testing sets. The training set will be used to train the machine learning model, and the testing set will be used to evaluate its performance.
5. Model selection: Select an appropriate machine learning algorithm that is best suited for detecting depression based on the nature of the data, such as decision trees, logistic regression, or support vector machines.
6. Model training: Train the selected machine learning algorithm on the training data set and optimize its hyperparameters to achieve the best performance.
7. Model evaluation: Evaluate the performance of the trained machine learning model on the testing data set using evaluation metrics such as accuracy, precision, recall, and F1 score.
8. Model interpretation: Interpret the trained machine learning model to gain insights into the factors that contribute to depression detection.
9. Model deployment: Deploy the trained machine learning model in a real-world setting to assist healthcare professionals in diagnosing and treating depression.

IV. FUTURE WORK

Depression, an issue that plagues many individuals everywhere, is undoubtedly a condition that severely affects countless people all over the world without any question whatsoever. The use of machine learning techniques for the detection and diagnosis of depression has been gaining traction in recent years. Still, there is a great deal of space for development and more research in this area. One potential direction for future work is the conduct of large-scale longitudinal studies. Such studies can involve collecting data from a large number of individuals over an extended period, allowing researchers to identify patterns and trends that may not be apparent in smaller studies. This can help to improve the accuracy and reliability of depression detection models, making them more useful in clinical practice. Another area of future work is multimodal data analysis. Integrating different sources of data such as text, speech, and physiological signals can enhance the accuracy of depression detection models. For example, analyzing speech patterns, facial expressions, and other physiological signals can provide a more comprehensive understanding of a person’s mental state. This can help healthcare providers to make more accurate diagnoses and provide more effective treatments. Personalized treatment is another area of future work in depression detection using machine learning. Machine learning models can be used to identify individuals who are at high risk of developing depression or who have a particular subtype of depression. This information can be used to personalize treatment plans, allowing healthcare providers to provide more effective and targeted

interventions. This can lead to improved outcomes for patients and better allocation of healthcare resources.

Incorporating social determinants of health is another important area of future work. Social determinants of health such as socioeconomic status, race, and ethnicity can significantly impact a person's mental health. Future work in depression detection using machine learning should consider integrating these factors into models to better understand and address health disparities. This can help to ensure that healthcare resources are allocated fairly and that all patients receive the care they need. Finally, explainable AI is an essential area of future work in depression detection using machine learning. One significant limitation of machine learning models is their lack of interpretability. To increase their clinical usefulness, it is essential to develop models that are transparent and explainable, allowing healthcare providers to understand how the model makes its decisions and trust its recommendations. This can help to ensure that these models are used ethically, responsibly, and with appropriate oversight.

In conclusion, future work in depression detection using machine learning has the potential to revolutionize mental health care. But it's imperative to address moral issues and make sure these models are created and applied in an ethically sound and open way. By doing so, we can help to improve the accuracy of depression detection models and provide better care to patients who need it most.

V. CONCLUSION

However, there are also significant ethical concerns that need to be addressed. Machine learning models can perpetuate existing biases and amplify existing health disparities. Therefore, it is essential to develop models that are transparent, explainable, and fair. Despite these concerns, the future of depression detection using machine learning is bright. There are numerous areas of future work that can help to improve the accuracy and usefulness of these models, such as large-scale longitudinal studies, multimodal data analysis, personalized treatment plans, and incorporating social determinants of health. We can enhance mental health treatment and make it more accessible and equitable for everyone by carrying out further research and development on these models. Various feature datasets are investigated and various machine learning techniques are used. Numerous preprocessing procedures are carried out, such as feature extraction and selection, data labelling, and alignment and preparation of the data. The SVM model transforms an enormously nonlinear classification issue into one that is linearly separable, achieving the best accuracy metric combinations. The DT model is thorough and has clear processes, but it can malfunction when presented with fresh data.

This study might be seen as a first step towards developing a comprehensive social media-based platform that analyses, forecasts, and suggests treatments for people experiencing mental and psychological problems.

The present work may be expanded in the future by examining more machine learning models that are unlikely to overfit the 15 data set, and by identifying a more reliable method for estimating the influence of the features.

REFERENCES

[1] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena, and V. Kumar, "A machine learning based depression analysis and suicidal ideation detection system using questionnaires and twitter," in 2019 IEEE Students Conference on Engineering and Systems (SCES), 2019, pp. 1–6.

[2] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, "A textualbased featuring approach for depression detection using machine learning classifiers and social media texts," Comput Biol Med, vol. 135, p. 104499, 2021.

[3] P. V. Narayanrao and P. L. S. Kumari, "Analysis of machine learning algorithms for predicting depression," in 2020 international conference on computer science, engineering and applications (iccsea), 2020, pp. 1–4.

[4] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," in 2017 international conference on information and communication technology convergence (ICTC), 2017, pp. 138–140.

[5] S. Paul, S. K. Jandhyala, and T. Basu, "Early Detection of Signs of Anorexia and Depression Over Social Media using Effective Machine Learning Frameworks.," in CLEF (Working notes), 2018.

[6] I. Fatima, B. U. D. Abbasi, S. Khan, M. Al-Saeed, H. F. Ahmad, and R. Mumtaz, "Prediction of postpartum depression using machine learning techniques from social media text," Expert Syst, vol. 36, no. 4, p. e12409, 2019.

[7] H. S. AlSagri and M. Ykhlef, "Machine learning-based approach for depression detection in twitter using content and activity features," IEICE Trans Inf Syst, vol. 103, no. 8, pp. 1825– 1832, 2020.

[8] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in reddit social media forum," IEEE Access, vol. 7, pp. 44883–44893, 2019.

[9] K. A. L. Govindasamy and N. Palanichamy, "Depression detection using machine learning techniques on twitter data," in 2021 5th international conference on intelligent computing and control systems (ICICCS), 2021, pp. 960–966.

[10] M. 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," Health Inf Sci Syst, vol. 6, no. 1, pp. 1–12, 2018.

[11] J. Liu and M. Shi, "A Hybrid Feature Selection and Ensemble Approach to Identify Depressed Users in Online Social Media," Front Psychol, vol. 12, no. January, pp. 1–13, 2022, doi: 10.3389/fpsyg.2021.802821.

[12] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gomez, "A text classification framework for ´ simple and effective early depression detection over social media streams," Expert Syst Appl, vol.

[13] D. P. Chandra and S. Rajarajeswari, "An optimized hybrid neural network model for detecting depression among twitter users," International Journal of Innovative Technology and ExploringEngineering, vol. 8, no. 10, pp. 2781–2795, 2019, doi: 10.35940/ijitee.J9590.0881019.

[14] P. Jain, K. Ram Srinivas, and A. Vichare, "Depression and Suicide Analysis Using Machine Learning and NLP," J Phys Conf Ser, vol. 2161, no. 1, p. 012034, 2022, doi: 10.1088/1742-6596/2161/1/012034

[15] R. Chatterjee, R. K. Gupta, and B. Gupta, "Depression detection from social media posts using multinomial naive theorem," in IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, p. 012095, 2021, doi: 10.1088/1757-899X/1022/1/012095.

[16] R. R. Baheti and S. Kinariwala, "Detection and analysis of stress using machine learning techniques," Int J Eng Adv Technol (IJEAT), vol. 9, no. 1, 2019.

[17] A. Trifan, R. Antunes, S. Matos, and J. L. Oliveira, "Understanding depression from psycholinguistic patterns in social media texts," in European Conference on Information Retrieval, 2020, pp. 402–409.

[18] D. B. Victor, J. Kawsher, C. Science, and E. Discipline, "Machine Learning Techniques for Study on Bengali Community," International Conference on Electronics, Communication and Aerospace Technology, 2020, pp. 1118–1126

[19] H. AlSagri and M. Ykhlef, “Quantifying feature importance for detecting depression using random forest,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 5, 2020.

[20] V. Leiva and A. Freire, “Towards suicide prevention: early detection of depression on social media,” in International Conference on Internet Science, 2017, pp. 428–436.