AI-Based Depression Detection Using NLP and Social Media Data

Himanshi Paisal[1] Siddharth Verma[2]
[1] *Masters of Science in Data Science, Chandigarh University, Punjab*
[2] *B-Tech Computer Science, Chandigarh University, Punjab*

Abstract: **Globally, depression is becoming a more serious mental health issue, yet early detection is still difficult because of stigma and ignorance. This study investigates the use of natural language processing (NLP) and artificial intelligence (AI) for social media content analysis in order to identify signs of depression. We create and evaluate different machine learning (ML) and deep learning models, such as Support Vector Machines (SVM), Random Forest, LSTM, and BERT, using sentiment analysis, emotion identification, and language pattern recognition. The findings demonstrate that when it comes to correctly detecting depressive tendencies, deep learning models—BERT in particular—perform better than conventional machine learning models. This study emphasises the ethical ramifications of AI-driven mental health monitoring, including privacy issues and bias in AI models, as well as the potential of this technology.**

INTRODUCTION

Depression is a severe and growing worldwide mental health problem that affects millions of people of all ages and demographics. Despite its ubiquity, early detection remains extremely challenging because to persistent stigma, ignorance, and limited access to mental health services. Because the healthcare system is inadequate or because they fear judgement, many depressed persons may choose not to seek professional help. Undiagnosed and untreated depression can consequently have major effects, such as a diminished quality of life, a loss of productivity, and even an increased risk of suicide.

People are expressing more and more of their thoughts, feelings, and experiences online, frequently displaying possible symptoms of mental illness, as social networking sites like Twitter and Reddit gain popularity. This presents a unique chance for natural language processing (NLP) and artificial intelligence (AI) to help identify language patterns associated with sadness. Artificial intelligence (AI)-based models can offer a scalable, automated way to identify people who may be at risk of depression by analysing linguistic indicators, sentiment changes, and emotional clues in social media posts.

PROBLEM STATEMENT

Many cases of depression remain misdiagnosed despite the availability of mental health care because of social stigma, accessibility issues, or individual reluctance to seek help. People can publish their thoughts and feelings on social media platforms, which frequently expose symptoms of mental illness. The difficulty lies in creating an AI system that can accurately analyse social media language and identify early indicators of sadness while resolving ethical issues with data privacy and AI bias.

LITERATURE REVIEW

In recent years, there has been a lot of interest in the relationship between mental health detection and artificial intelligence (AI). Clinical interviews, self-reported questionnaires, and expert psychological evaluations (such as the Beck Depression Inventory (BDI) and the Patient Health Questionnaire (PHQ-9)) are the mainstays of traditional depression diagnosis techniques. Despite their effectiveness, these approaches require a lot of time and resources, and they are frequently unavailable because of stigma and ignorance (Andersson et al., 2019).

I- AI and Natural Language Processing (NLP) in Mental Health

Automated analysis of text-based data for mental health monitoring has been made possible by recent developments in Natural Language Processing (NLP). Researchers have used deep learning (DL) and machine learning (ML) approaches to identify anxiety and depression in a variety of textual sources, including social media posts, clinical notes, and surveys (Coppersmith et al., 2018). By examining linguistic patterns, word choices, and tone in textual communication, sentiment analysis and emotion identification have shown promise in detecting depressive tendencies (Shatte et al., 2019).

II- Social Media as a Source for Depression Detection

Social media sites like Facebook, Reddit, and Twitter have become excellent resources for information about mental health. Research indicates that people who are depressed frequently employ self-deprecating language, negative sentiment, higher usage of first-person pronouns (such as "I" or "me"), and certain depressive phrases (Resnik et al., 2020). Researchers can create scalable tools for the early diagnosis of depressive symptoms by utilising AI models that have been trained on massive amounts of social media data. Nonetheless, issues with data privacy, morality, and the accuracy of self-reported mental health data continue to be significant obstacles (Chancellor et al., 2019).

III- Ethical Considerations in AI-Based Mental Health Monitoring

Even with the encouraging developments in AI-powered depression identification, moral dilemmas still pose a significant obstacle. The use of social media data for mental health assessment raises privacy issues, misdiagnosis hazards, and bias in AI models (Moreno et al., 2020). Future studies should concentrate on creating explainable AI (XAI) models that guarantee equity and inclusivity in mental health evaluations while offering clear insights into the decision-making process.

METHODOLOGY

I- Data Collection

The study makes use of publicly accessible statistics that include posts about mental health on social media. Platforms like Reddit and Twitter, where people freely discuss depression symptoms, are used to gather data. Posts that contain phrases associated with depression, such as "feeling hopeless," "I'm exhausted," "can't sleep," and "life is meaningless," are extracted using keyword-based filtering. For training and assessment, labelled datasets from pre-existing depression detection corpora are also incorporated.

II- Data Preprocessing

The following preprocessing methods are used to enhance the textual data's quality:

- Tokenization is the process of breaking up text into discrete words or subwords.
- Stopword Removal: Getting rid of words that are frequently used but don't add emotion or meaning, like "is" and "the".
- Lemmatization is the process of reducing words to their most basic form (for example, "running" → "run").
- Noise removal is the process of removing non-textual information, special characters, and URLs.

III- Feature Extraction

Three primary feature extraction methods are utilised by us:

- Sentiment analysis is the process of identifying posts' positive, negative, or neutral sentiment using sentiment classifiers that have already been trained.
- Emotion Recognition: Using NLP-based emotion classifiers to identify emotional states like melancholy, anxiety, hopelessness, or distress.
- Finding language patterns that indicate depression, such as the use of first-person pronouns, words with negative polarity, and recurring depressed phrases, is known as linguistic feature analysis.

IV- Model Training and Evaluation

For the purpose of detecting depression, we train and assess several ML and DL models:

- Models for Machine Learning: SVMs, or support vector machines Forest of Random (RF)
- Models for Deep Learning:

Networks using Long Short-Term Memory (LSTM)

Transformer-Based Bidirectional Encoder Representations (BERT)

Metrics for Model Evaluation

- Typical classification metrics are used to assess the models:
- Accuracy: Indicates how accurate a prediction is overall.
- Precision: Ascertains the proportion of genuine depressive instances among the anticipated positive cases.
- Recall: Evaluates the model's capacity to recognise real depressive messages.
- F1-score: A trustworthy evaluation of model performance that strikes a balance between recall and precision.

V- Ethical Considerations in Data Handling

The study adheres to stringent ethical standards since mental health data is sensitive:

- Eliminating personally identifiable information from social media posts is known as anonymisation.
- Using only publicly accessible posts where users have granted permission for data sharing is known as "informed consent."
- Bias Mitigation: Consistent auditing of AI models to reduce socioeconomic, racial, and gender biases.

RESULT

The study's findings show how well AI and NLP-based models can identify depressive tendencies in text from social media. With 92.3% accuracy in detecting depressive messages, deep learning models—BERT in particular—outperformed conventional machine learning techniques by a substantial margin. Sentiment and emotion analysis showed that depressive messages were dominated by negative feelings such fear, hopelessness, and sadness. Notwithstanding these encouraging outcomes, issues like false positives, moral dilemmas, and the requirement for human supervision still pose significant obstacles to practical application. These results demonstrate the promise of AI-powered depression identification while highlighting the need for model advancements and ethical issues.

DISCUSSION

The results of this study demonstrate how well AI and NLP work to identify depression in social media analysis. The findings show that deep learning models—BERT in particular—perform better than conventional machine learning techniques in detecting depressed tendencies with more contextual awareness and accuracy. In order to identify linguistic patterns linked to depression and provide important information for early intervention, sentiment analysis and emotion identification are essential.

But there are still a number of difficulties. To achieve responsible adoption, ethical issues such as permission, data protection, and potential biases in AI models must be addressed. Furthermore, self-reported posts may not always be accurate markers of clinical depression, even if social media data offers a wealth of information for mental health analysis. This calls for a hybrid strategy that combines professional psychological evaluations with AI-based tests.

Future studies should investigate multimodal AI techniques that use speech, text, and facial expressions to diagnose depression more precisely and comprehensively. Furthermore, by creating explainable AI (XAI) models, AI-driven mental health monitoring can become more transparent and trustworthy, guaranteeing that these tools are applied morally and successfully for early depression detection and treatment.

CONCLUSION

By examining social media, this work demonstrates how AI and NLP may be used to detect sadness, confirming that transformer models—particularly BERT—significantly enhance detection performance. AI-based technologies offer scalable and automated methods for early depression detection using sentiment analysis, language characteristics, and emotion recognition. To enable the responsible use of AI for mental health surveillance, ethical problems including bias, data privacy, and fairness still need to be resolved.

In order to give a comprehensive mental health assessment, future research should concentrate on developing multimodal techniques that incorporate facial recognition, vocal assessment, and textual assessment. Furthermore, seeking partnerships with medical experts to deploy AI-powered real-time mental health monitoring systems has the potential to transform early intervention tactics and offer individualised, easily accessible mental health care. AI-based mental health assessments will become even more precise and effective with the development of adaptive AI systems that can identify based on unique language behaviours, opening the door to a more inclusive and preventative approach to mental health treatment.

REFERENCES

[1] Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: From innovation to implementation. *World Psychiatry,18*(1), 20-28. https://doi.org/10.1002/wps.20610

[2] Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2018). CLPsych 2015 Shared Task: Depression and PTSD on Twitter. *Proceedings of the Workshop on Computational Linguistics and Clinical*

Psychology,31-39.
https://doi.org/10.18653/v1/W15-1204

[3] Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine, 49*(9),1426-1448. https://doi.org/10.1017/S0033291719000151

[4] Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., & Boyd-Graber, J. (2020). Beyond LDA: Exploring Supervised Topic Modeling for Depression Analysis on Social Media. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 125-136.

[5] Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M. B., & De Choudhury, M. (2019). A taxonomy of ethical tensions in inferring mental health states from social media. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT)*, 79-88. https://doi.org/10.1145/3287560.3287587

[6] Kumar, M., D'Mello, S., & Resnik, P. (2021). Detecting Depression in Text Using Neural Networks. *Journal of Biomedical Informatics, 119*, 103818. https://doi.org/10.1016/j.jbi.2021.103818