**RESEARCH GAPS**

1. -There is relatively more accurate results but still the error analysis shows some that the results can be improved . This model is trained on a specific dataset of a hospital which don’t generalizes the total population. And the testing was done with internal data not providing its capabilities in different settings . Incorporating a more set of features, such as genetic information, socioeconomic factors, or detailed behavioral assessments, could enhance predictive accuracy.This research involves the text based dataset which disregards the depression indicators that can be found in images and audio .
2. -This study's Reddit-based approach faces key limitations compared to Twitter-based depression detection. While Reddit offers structured, long-form posts, Twitter provides noisier but more linguistically diverse real-time data with self-reported labels, reducing ambiguity. The reliance on TF-IDF lacks depth compared to advanced embeddings (BERT, word2vec) used in Twitter studies, and the lower F1-score for depression (0.67 vs. anxiety’s 0.84) highlights class imbalance issues.  The system worked worse for depression than anxiety, and didn't use important clues like how often people post or what emojis they use. Unlike Twitter research, this work didn't look at how often people post and it also didn't address potential biases. Future work should integrate contextual embeddings, multimodal data.
3. -Existing 2022 system developed a hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM *focusing on using English text only, excluding emojis, videos, and foreign languages, with noisy data cleaned through pre-processing.What we are proposing in 2025* is *to develop a Deep Learning Approach for Depression Classification and Prediction from User Tweets and social media focusing not only on using English text but also including videos, and foreign languages*
4. - Even though it considered using Reddit and Twitter dataset but still it cant be robust to any other environment as it is data centric to reddit and twitter .This study also just collaborated with text data not including images and audio files .
5. - Earlier studies used only one type of data like surveys. This project is different because it mixes many types of data—survey answers, social media activity, and health info. This gives better and more accurate results.
6. -The research gaps in the paper include the use of a small and imbalanced dataset (DAIC-WOZ), limiting generalizability and performance. The models lack advanced contextual understanding for text and broader integration of multimodal features. Real-world scalability and robustness against variability in voices and noise are not addressed. Additionally, the focus on binary classification misses opportunities to detect depression severity or temporal progression.
7. -Research Gap and Improvement In previous research, single deep learning models were used for health prediction, which had limitations in handling complex and sequential healthcare data. This project addresses that gap by introducing a hybrid CNN-LSTM architecture. CNNs are efficient in feature extraction from sensor data, while LSTMs are effective in processing time-series data. By combining these models, the system achieves higher accuracy and robustness in predicting patient conditions. This novel approach represents a significant step forward in smart healthcare technologies
8. -Problems in This Project and Potential Improvements The authors do imply that there is a problem that can be improved. Here are a few areas where the authors suggest potential improvements or where limitations of their study suggest room for improvement:

● Further Optimization of ESG for Low-Powered Devices: The authors specifically mention that Electra Small Generator (ESG) shows promise, but they propose "further optimization of ESG to make it suitable for low-powered devices". This suggests that while ESG is effective, it may not be ideal for resource-constrained applications in its current state.

● Generalizability: While the study achieves strong results on their specific Twitter dataset, the generalizability of these models to other social media platforms or different datasets could be explored further. Different social media platforms may have unique language styles and user behaviors, which could impact the models' performance.

● Dataset Bias: The authors used a dataset created from tweets with depression-related hashtags. Datasets created in such a way can introduce bias. Hashtags may not perfectly capture all instances of depression-related expression, and the sentiment analysis tools used for labeling (VADER, TextBlob) might have limitations. Improving the dataset creation process, perhaps with more diverse data sources or more refined labeling techniques, could be beneficial.

● Real-time Application: The study focuses on offline analysis of tweets. There could be further work on how to implement these models in real-time systems for immediate depression detection and intervention.

COMMON ANALYSIS :

OBJECTIVES :

PROBLEM STATEMENT :