ISSN: 2583-9292 (ONLINE)
DOI: 10.xx/ijdsml.2025.xx

SUICIDAL POST DETECTION ON REDDIT USING DEEP LEARNING TECHNIQUES

Swati Bansode, Vaishali Hirlekar and Shashikant Radke
Department of Computer Engineering, Shah and Anchor Kutchhi Engineering, India

Abstract

Suicide is a leading cause of death worldwide, particularly among the young generation. The increase in suicidal posts on social media platforms such as Reddit has presented both challenges and opportunities for mental health intervention. Our work aims to use vast data generated by individuals on Reddit by using advanced deep learning techniques to identify suicidal posts and non-suicidal posts. The dataset used is collected from two different datasets: one dataset curated from the Reddit API called (PRAW) with the help of various subreddits (e.g., SuicideWatch, Anxiety, and Depression) and neutral topics (e.g., Jokes, Movies, Popular, Books) and another publicly available dataset from Kaggle. The proposed model uses long-short-term memory LSTM, bidirectional LSTM (Bi-LSTM), gated recurrent units (GRU), bidirectional GRU (Bi-GRU), and modified BERT-based transformers. The BERT-based model performed better in both datasets compared to other models with an accuracy of 98%, a precision of 98. 5%, and a recall of 98. 5%. These experimental results successfully verify the theoretical efficiency and adaptability of the proposed model in real-time suicidal post-detection.

Keywords:

Suicidal Post-Detection, Deep Learning, Natural Language Processing, Reddit, BERT, LSTM, Bi-LSTM, GRU and Bi-GRU

1. INTRODUCTION

One of the biggest issues facing our age is suicide. Every year, 76,000 people harm themselves and many more attempt suicide. Every suicide is a tragedy that has a lifelong impact on survivors and affects families, communities, and even nations. Suicide affects people of all ages and was the third most common cause of death worldwide in 2021 among those aged 15 to 29 years [1]

Since social networking websites have evolved into a platform for individuals to express their emotions, such as tension, anxieties, sadness, and suicidal thoughts, it is now crucial to detect suicidal thoughts on these platforms. With the rise of social media sites like Reddit, deep learning and natural language processing (NLP) have demonstrated considerable promise in analyzing content generated by users to identify suicidal thoughts and assist individuals with mental health difficulties.

By using advanced deep learning models such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Gated Recurrent Units (GRU), Bidirectional GRU (Bi-GRU), and a customized BERT-based transformer model, our model's main objective is to identify suicidal posts on the Reddit website.

The Reddit Dataset is used to train both models, which the Python Reddit API Wrapper (PRAW) [2] was used to gather. The second dataset used in this work is a publicly available dataset from Kaggle. Posts from a variety of Reddit communities, or subreddits, such as SuicideWatch, Anxiety, and Depression, as well as general interest subreddits like Books, Movies, Popularity, and Jokes, make up our dataset. This varied dataset provides a balance for model training and model analysis by capturing not only posts on mental health but also natural and non-suicidal posts.

This study's contributions are threefold:

- **Knowledge Discovery**: This study analyzes linguistic and behavioral patterns suggestive of suicide ideation in online user material using deep learning techniques. Suicidal people often use personal pronouns, convey negativity, and explain their suffering in the present tense, while referring to their hopelessness and suicide intentions in the future tense, as each post reveals.
- **Dataset and Platform**: Our work uses a dataset that was curated from Reddit using the PRAW Reddit API, which included posts from neutral post from subreddits (Books, Movies, Popular, and Jokes) and subreddits like SuicideWatch, Anxiety, and Depression. This dataset can be applied to general communication patterns in various circumstances as well as suicide post detection.
- **Deep Learning Models and Benchmarking**: The performance of different deep learning methods, such as LSTM, Bi-LSTM, GRU, Bi-GRU, and a customized BERT-based model, is compared in our work. Linguistic, contextual, and topic-specific information are used to instruct and evaluate these models. We also used the publicly accessible dataset to compare our results. Future research in social media mental health detection can build on the baseline results, which can be used to detect suicide posts.

This work intends to provide reliable and scalable methods for identifying suicide posts on Reddit by fusing advanced methods of deep learning with a dataset. The findings could support real-time mental health monitoring and interventions to help identify those who are at risk early.

2. RELATED WORK OF EXISTING METHODS

With a focus on Formspring.me, where anonymity promotes a large amount of bullying content, Reynolds et al. [3] used machine learning models to identify cyberbullying on social media platforms. They demonstrated the viability of automated detection methods in reducing online harassment by achieving an accuracy of 78.5% using labeled datasets and classification algorithms like C4.5 decision trees.

Transformer-based models for identifying suicidal thoughts in social media posts, specifically from Reddit's SuicideWatch community, were investigated by Haque et al. [4]. After putting the BERT, ALBERT, RoBERTa, and XLNet models into practice, they discovered that transformer-based architecture performed noticeably better than traditional deep learning techniques like Bi-LSTM. RoBERTa was the most accurate of these models, demonstrating its ability to analyze textual signs of mental health concern.

This work was expanded to Twitter by Ananthakrishnan et al. [5], who used BERT-based transformer models, BERT, DistilBERT, ALBERT, RoBERTa, and DistilRoBERTa, to study suicidal intention detection. According to their findings, RoBERTa performed better than expected, achieving 95.39% testing accuracy and 99.23% training accuracy. Their research highlights how well pre-trained transformer models can detect mental health issues in text from social media and points to the potential for future improvements using graph transformers and quantum machine learning.

A multi-class machine learning classifier was created by Rabani et al. [6] to recognize various degrees of suicidal risk in social media messages. They identified features from a dataset gathered from Reddit and Twitter using an Enhanced Feature Engineering Approach for Suicidal Risk Identification (EFASRI). SVM, Random Forest, and Extreme Gradient Boosting (XGB) were used in their investigation; XGB had the highest accuracy of 96.33%, highlighting the significance of feature selection in raising prediction accuracy.

In order to detect suicidal thoughts in online user content, Ji et al. [7] investigated supervised learning approaches. Their study extracted linguistic, syntactic, and topic-based elements from Twitter and Reddit's SuicideWatch data. They contrasted a number of classifiers, including deep learning techniques like LSTM and XGBoost as well as conventional machine learning models. According to their findings, linguistic cues, like the use of personal pronouns and negative emotional expressions, are essential for identifying people who are suicidal.

Using a variety of linguistic, topic-based, and sentiment-related variables, Chatterjee et al. [8] examined data from Twitter. Their research showed that feature selection is useful in enhancing classification performance, as seen by the 87% accuracy of a Logistic Regression model.

Using LSTM and Random Forest classifiers on data gathered from SuicideWatch and Depression subreddits via the Pushshift API, Imam et al. [9] looked at the detection of suicidal thoughts on Reddit. With a 93% accuracy rate, their method demonstrated the potential of deep learning models for suicide-related material classification.

Jindal and Malhotra [10] carried out the classified methods according to supervised, unsupervised, and semi-supervised deep learning algorithms after reviewing 96 research publications. The study's findings support the use of deep learning in mental health surveillance by showing that transformer-based models such as BERT and XLNet perform better than conventional machine learning methods.

Support Vector Machine (SVM), Logistic Regression, and AdaBoost were used by Chadha et al. [11] to identify suicidal ideation in social media data, with a primary focus on Reddit posts. Their research showed that TF-IDF embedding and machine learning models are successful in detecting suicidal behavior, with SVM achieving the highest precision (80.72%) and Logistic Regression achieving the best accuracy (80.75%) and recall (77.81%).

In order to detect suicidal ideas in posts from Kenyan social media platforms such as Facebook and Twitter, Andrew Kipkebut and Emmanuel Chesire [12] used BERT in their study. Following extensive preprocessing and tokenization, the BERT-based model achieved high precision and recall scores (91 and 87, respectively).

In their work, Jaideep Yadav et al. [13] developed a BERT-based approach for identifying cyberbullying on social media, with an emphasis on text from sites such as Wikipedia and Formspring. To categorize bullying content, the study used a single linear neural network layer and the pre-trained BERT model. Since of its transformer design, BERT is more accurate than standard models since it can capture context inside text. They obtained a high F1-score of 94% by oversampling the minority (bullying) class in the Formspring dataset.

3. PROPOSED MODEL FOR SUICIDAL POST DETECTION

Table.1. Previous Work Comparison

Authors	Algorithms/ Models	Dataset	Performance
Reynolds et al. [3]	C4.5 Decision Tree, Instance-Based Learner, WEKA Toolkit	Interviews with 47 suicide attempt survivors in Pune, India	Both a C4.5 decision tree learner and an instance-based learner were able with 78.5% accuracy
Haque et al. [4]	Bi-LSTM, Transformers (BERT, ALBERT, RoBERTa, XLNet)	Reddit (SuicideWatch)	Transformer models ROBERTa achieved 95.21% accuracy
Gayathri Ananthakrishnan et al. [5]	BERT, DistilBERT, ALBERT, RoBERTa, DistilRoBERTa.	Twitter dataset with 9119 tweets	RoBERTa achieved 95.39% accuracy on test data
Syed Tangelo Rabani, et al.[6]	SVM, Random Forest (RF), Extreme Gradient Boosting (XGB)	Twitter and Reddit dataset	XGB achieved an accuracy of 96.33%
Shaoxiong Ji, et al.[7]	SVM, Random Forest, GBDT, XGBoost, LSTM, Word Embedding	Reddit SuicideWatch, Twitter	LSTM 97.86% highest accuracy
Moumita Chatterjee et al.[8]	Logistic Regression, Sentiment Analysis, LDA	188,704 tweets	Logistic Regression accuracy of 87%
MD. Rafi Imam et al.[9]	LSTM, Random Forest	Reddit data	Achieved up to 93% accuracy with LSTM.
A. Malhotra and R. Jindal[10]	ANN, DNN, CNN, RNN, GRU, Bi-RNN, Bi-LSTM, Bi-GRU, BERT, XLNet, GloVe, fastText, ERDE (Early Risk Detection Error)	Twitter, Reddit, CLPsych, eRisk	BERT, demonstrated an highest accuracy
Akshma Chadha, et al.[11]	SVM, Logistic Regression, AdaBoost, TF-IDF	Reddit (SuicideWatch)	Logistic Regression achieved highest accuracy (80.75%)
Andrew Kipkebut et al.[12]	BERT, NLP preprocessing, Tokenization	Social media platforms in Kenya, including Twitter and Facebook	BERT Achieved 91% precision

Jaideep Yadav et al. [13]	BERT model, Transformer architecture	Formspring (QandA) and Wikipedia talk pages	Achieved up to 98% accuracy on Formspring, 96% on Wikipedia

Table.2. Sample suicidal posts on Reddit

Suicidal Posts	**Type of post (subreddit)**
Going to kill myself in 8 hours I have decided on a method already, I just need to write a suicide note and I will be done. Not sure why I wrote this but I guess I just wanted to tell somebody Not changing my mind	r/SuicideWatch
I just wanna to rest in peace forever but dk when. Death is invitable for everyone and so for me. I just can't wait to die anymore but I wonder when I'll kms and finally will leave my shitty body and this planet and people. I want to feel how will. I feel when I'll not be breathing and this cursed heart will not be beating. I can't believe I failed in this life wow! Can't wait to leave but when?	r/SuicideWatch
Don't want to live. Too scared to die. I've had suicidal ideations since I was 16 yo (I'm 36 now). My life is chaotic and looks about how you'd expect from someone who is basically a half life , walking around wishing I were dead.	r/SuicideWatch
I deserve to die. I'm a loser. A failure. A dumb ass. I want to end it all. I'm sure my family cares. But I don't want to disappoint them. I've gotten a lot of hate from people. I suck at every minuscule thing. I've asked for help multiple times but people ignore me	r/depression
Antenatal depression I feel done. Very done. I don't have it in me to care about anything anymore. I'm 25 weeks pregnant and been let down by everyone supposed to be my care team to the point I don't want to be pregnant. (I chose to keep my baby in the first place and wanted him. I'm in the UK and had access to care if I had chosen not to) I tried to stab myself in the stomach this morning because things got so bad after trying to stab my wrist. My partner stopped me and I didn't manage to actually do anything.	r/depression
Want to kill myself because of my height.	r/Anxiety

Table.3. Rules and Example of suicidal and Non-suicidal Text separated by community

Categories	***Community***	***Rules***	***Example***
Suicidal Text	r/Suicide Watch r/Anxiety r/depression	Suicidal related Thought	I will kill myself. I can't do it anymore.
		Suicidal related thought. Potential suicidal thought.	I think about suicide every day.
Non-Suicidal Text	r/Jokes	Discussing about death	I was choking on a piece of food and it kind of scared some people. I almost died.
	r/movies	Discussing others suicide	Japanese Actress Sei Ashina Dies of Suicide at Age 36.
	r/books	Not related to suicide	is anyone else hyped about the new Ron Cher now biography on Mark Twain?
	r/popular	Not related to suicide	The US has the fourth highest suicide rate

3.1 DATA COLLECTION AND EXPLORATION

The data was collected from the social media platform Reddit through their APIs.

3.1.1 Reddit Dataset 1:

Reddit's extensive variety of user-generated content, which includes posts, comments, and debates on a wide range of subjects, made it the study's main source of data. The Python Reddit API Wrapper (PRAW), a Python framework for secure interaction with the Reddit API, was used to retrieve the data. Reddit's API requires authentication in order to access the data. Creating an application on Reddit's developer site and acquiring the client id, client secret, password, user agent, and username were the steps in this process.

3.1.2 Dataset Description:

Posts from subreddits with text fields such titles, body content, and unique identifiers to protect user anonymity created to the gathered dataset. Based on their intent, posts were divided into two categories: suicidal and non-suicidal. Some of the example of posts are shown in Table.3. The following subreddits were the source of posts with suicide intent: Depression, Anxiety, and SuicideWatch. The following sources provided posts classified as non-suicidal: r/Jokes, r/Movies, r/Books, and r/Popular

Fig.1. Keyword Based Word Cloud

3.1.3 Data Labeling:

A keyword-based search strategy was used to distinguish between posts that were suicidal and those that weren't. Posts with suicidal intent were identified using keywords that are suggestive of suicidal ideation, such as "disappear", "die", "worthless" and "pain" A word cloud was used to illustrate the keyword identification procedure Fig.1. According to this analysis, there were a lot of phrases associated with mental health issues and hopelessness. Some of the examples are shown in Table.2.

3.1.4 Dataset Statistics and Imbalance:

The final dataset contained a total of 8,237 posts in all, which split into two groups 5,278 non-suicidal posts: 2,959 suicidal posts. With more non-suicidal posts than suicidal ones, this distribution showed an imbalance in the dataset. Oversampling was used to improve model training performance and balance the dataset to solve this problem.

3.1.5 Exploratory Analysis:

To learn more about the terms that people posting on the site use most frequently, the data was further examined. The recurrent themes of emotional suffering, uncertainty, and mental health issues were emphasized using a word cloud Fig.2(a)and(b) of frequently used terms including "medication", "choice", "disorder" and "regret". The platform's function as a forum for people to express their feelings and look for support or validation was highlighted by this analysis.

3.1.6 Public available Dataset 2:

A publicly accessible dataset from Kaggle was used to compare the first dataset. There were 232,079 posts in all in this dataset, which was likewise taken from Reddit. A more thorough grasp of the trends and patterns connected to Reddit user-generated material was made possible by the inclusion of this larger dataset, which broadened the analysis's impact.

3.1.7 Word Cloud:

Words like "feel", "want", "life", "people", and "time" predominate in the word cloud in Fig.2b, highlighting existential issues and emotional expressiveness. With frequent references to words like "friend", "make" and "doing", words like "die", "lonely" and "happy", imply conversations focused on emotional health and interpersonal relationships. With "know", "want", "think", "life" and "time" as the most common terms, the Fig.2(a) and Fig.2(b) word cloud, on the other hand, shows a noticeable change. The increase in the use of words like "know", "think" and "reason" indicates a higher emphasis on cognitive and analytical conversations. The Fig.2(a) word cloud shows a emphasis on emotional vocabulary like "feel" and "die" and a greater emphasis on relational and action-oriented words like "help", "talk" and "friend" Both word clouds retain terms like "kill" and "end" indicating continuous references to extreme emotional states.

Fig.2(a). Word Cloud of Dataset 1

Fig.2(b). Word Cloud of Dataset 2

4. METHODOLOGY

The aim of this study is to determine the top-performing algorithms that assist in detecting suicidal posts on Reddit. The various deep learning algorithms, including LSTM, Bi-LSTM, GRU, Bi-GRU, and one transformer that we customized in the layer BERT model, have been compared in this study. Two datasets were employed in our work. Our suggested framework for this paper is shown in Fig.3.

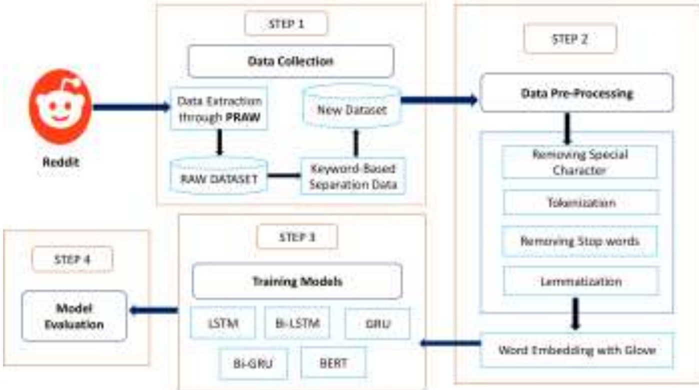


Fig.3. Proposed model

4.1 DATA PRE-PROCESSING AND FEATURE ENGINEERING

Due to its user-generated nature, the textual data gathered from Reddit is naturally noisy. It contains a number of inconsistent and unnecessary components, including slang, special characters, casual language, and excessive data. This data was subjected to a systematic pre-processing and feature engineering procedure to guarantee its suitability for computational analysis and deep learning models. The steps listed below describe the specific procedures used:

4.1.1 Removing Special Characters:

Special letters, symbols, and other non-alphanumeric components are frequently found in textual data, but they don't add to the text's semantic meaning. In this step, components including punctuation (.,,!,?), symbols (#, $,\@), emoji's and other irrelevant characters had to be found and eliminated.

For example:

- Original text: "I'm feeling #anxious! □□ Can't seem to calm down @ all…"
- After cleaning: "I'm feeling anxious Can't seem to calm down"

The dataset was cleaned up by eliminating these components, ensure that only significant text remained. During text processing, this procedure helps cut down on background noise and any distractions.

4.1.2 Tokenization:

The practice of dividing text into discrete words, phrases, or significant subunits known as tokens is known as tokenization. This process simplifies the data and makes word-level analysis easier. Natural language processing (NLP) libraries like NLTK were used for tokenization in order to appropriately handle linguistic subtleties.

For example:

- Text: "I am feeling overwhelmed today."
- Tokens: ["I," "am," "feeling," "overwhelmed," "today"].

For word-level operations like frequency analysis, stop word elimination, and lemmatization to be enabled, this stage is especially important. Tokenization also gets the data ready for vectorization methods like Word Embeddings.

4.1.3 Removing Stop Words:

Common English words known as "stop words" have minimal semantic significance when used in an analysis. The words "the", "is", "and", "to", "of" and "in" are among the examples. Although they are commonly used, these terms don't help individuals grasp the text's underlying meaning or intent.

These words were effectively eliminated from the dataset using a predetermined list of stop words from NLP libraries.

For example:

- Original tokens: ["I," "am," "feeling," "very," "overwhelmed," "today"]
- After stop word removal: ["feeling," "overwhelmed," "today"]

This stage improves the quality of features retrieved for deep learning by guaranteeing that the emphasis stays on the more significant and content-rich words.

4.1.4 Lemmatization:

A text normalization method called lemmatization breaks words down to their basic or root form. Lemmatization takes into account a word's grammatical context to ensure an accurate transformation, in contrast to stemming, which merely truncates words to a common prefix.

For example: "running", "ran" and "runs" → "run", "Better" → "good"

By treating different forms of the same term as a single item, this phase eliminates redundancy and improves the dataset's consistency. Lemmatization modules in NLP libraries are used in the process.

4.1.5 Word Embedding with GloVe:

A tool in Natural Language Processing (NLP), word embedding converts words into numerical representations so that text data can be processed by deep learning models. A popular technique for creating dense vector representations that capture the semantic links between words in a continuous vector space is called GloVe (Global Vectors for Word Representation).

Pre-trained GloVe embeddings were used for this study. These embeddings come in several dimensions, including 50, 100, 200, and 300-dimensional and Common Crawl. In these dimensions, each word is represented as a dense vector that maintains relationships between words. Vectors, and were trained on extensive corpora like Wikipedia

4.2 TRAINING MODELS

Batch processing is made possible by padding these tokenized sequences to ensure consistent length, which we describe in the previous step. In this stage, each word is converted into a multi-dimensional vector that captures its semantic meaning. The deep learning models use this embedding as their input for training models.

4.2.1 LSTM (Long Term Short Memory):

In order to handle sequential input, LSTM is made to learn both short-term and long-term dependencies. Word by word as it processes the sequence while preserving a concealed internal state that changes in response to new inputs. The LSTM generates a context vector that summarizes the sequence by carefully choosing whether information to keep, update, or delete using internal procedures called gates. A sigmoid activation function generates a probability indicating the presence of suicidal ideation once this context vector is sent to a fully connected (dense) layer.

4.2.2 Bi-LSTM (Bidirectional Long Term Short Memory):

By processing the sequence both forward and backward, the Bi-LSTM improves upon the LSTM. The backward LSTM records dependencies in reverse order, from the end to the beginning, whereas the forward LSTM records dependencies from the beginning to the end of the sequence. A richer representation of the text is produced by concatenating the outputs from both sides, which captures context from both past and future words. Because of this, Bi-LSTMs are especially good at detecting complex emotional meaning in post.

4.2.3 GRU (Gated Recurrent Unit):

It turns the input and forget gates into a single update gate, simplifying the LSTM architecture. This simplified architecture preserves the capacity to identify important connections in the text while lowering computing complexity. The sequence is efficiently processed by the GRU, which keeps relevant details while eliminating unnecessary ones. It produces a context vector that is sent to a dense layer for prediction, just as the LSTM.

4.2.4 Bi-GRU (Bidirectional Gated Recurrent Unit):

It adds bidirectional processing to the GRU, expanding on it. It has two GRU layers, just like the Bi-LSTM, one of which processes the sequence forward and the other backward. A context-aware representation of the sequence is produced by concatenating the outputs from both directions. This method is appropriate for situations where comprehending the entire context is essential since it combines the effectiveness of GRUs with the capacity to record bidirectional dependencies.

4.2.5 BERT (Bidirectional Encoder Representations from Transformers):

A transformer-based model that has been fine-tuned for its purpose to capture deep contextual embeddings has been pre-trained on extensive text corpora.

5. DATA PREPROCESSING

A pre-trained BERT tokenizer is used to prepare raw text data for model input at the beginning of the workflow. There are multiple tasks involved in this step:

- Tokenization: Words are mapped to their respective IDs in BERT's lexicon by converting raw text into tokenized sequences.
- Padding and Truncation: Longer sequences are truncated, and shorter ones are padded with zeros to standardize sequence length (e.g., 128 tokens).
- Attention Mask Generation: Masks are made to differentiate between padding tokens (0) and important tokens (1).

Compatibility with BERT's input criteria is ensured by the processed data, which is saved as input_ids and attention_mask tensors.

5.1 MODEL ARCHITECTURE

With a unique design made for binary classification, the model is based on BERT's encoder:

- BERT Encoder: uses self-attention methods to extract input token contextual embeddings.
- Classifier layer: To identify inputs (such as "suicide" or "non-suicide"), a softmax activation function is added to a fully connected network with dense layers.

5.1.1 Training:

During the training phase, supervised learning approaches are used to maximize the model's parameters:

- Loss Function: The error between the expected and actual labels is computed using CrossEntropyLoss.
- Optimizer: The Adam optimizer uses gradients calculated during back propagation to adjust model weights.
- Batch Processing: To increase processing efficiency, data is separated into batches. The model receives input tensors (input_ids and attention_mask) for every batch.

The loss is computed by comparing the generated predictions with the labels of the ground truth. Model weights are improved by backpropagating gradients.

To assess generalization and keep focus on overfitting, the model is validated after each period of training.

5.2 EVALUATION

To make sure the model performs well when applied to data, its performance is assessed both during and after training:

- **Loss and Accuracy Tracking**: To evaluate the model's learning curve and convergence, training and validation loss as well as validation accuracy are shown over epochs. As seen on Fig.6.
- **Overfitting Detection**: The difference between validation loss and training loss is tracked. To avoid overfitting, strategies like early halting and regularization are used.

5.3 RESULT

Five deep learning models, LSTM, Bi-LSTM, GRU, Bi-GRU, and BERT, were assessed for their ability to identify suicidal ideation in Reddit posts. Two datasets were employed for the evaluation: Dataset 1, which was a carefully selected collection from Reddit subreddits devoted to general issues and mental health, and Dataset 2, which is a publicly accessible Kaggle dataset with a wider range of Reddit postings. The outcomes show how BERT transformer-based architecture, which is excellent at capturing semantic and contextual details, allows it to perform better in both datasets.

BERT obtained the highest accuracy of 98.5% in Dataset 1, as indicated in Table.4, with matching precision, recall, and F1-score values of 98.5%. This result was marginally better than Bi-GRU (97.9% accuracy) and much better than LSTM (93.3% accuracy). The significance of utilizing forward and backward contextual dependencies in the classification task was underscored by the superior performance of the bidirectional models (Bi-LSTM and Bi-GRU) over their uni-directional counterparts (LSTM and GRU). With an accuracy of 97.4%, GRU outperformed LSTM by a small margin. This was probably because of its more straightforward architecture, which maximizes sequential learning.

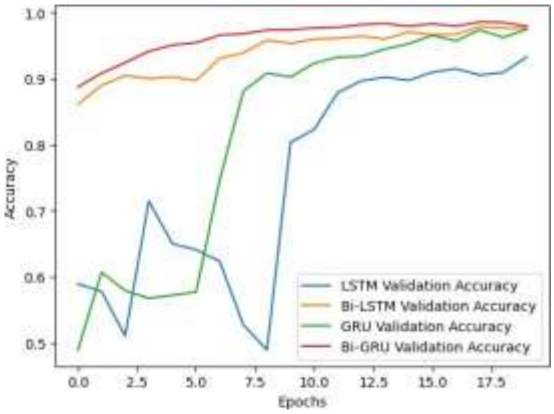
Table.4. Dataset 1

Models	Accuracy	Precision	Recall	F1 Score
LSTM	93.3	92.7	93.6	93.1
Bi-LSTM	97.6	96.2	97.6	97.6
GRU	97.4	98.2	96.5	97.3
Bi-GRU	97.9	96.3	96.8	97.9
BERT	98.1	98.5	98.2	98.5

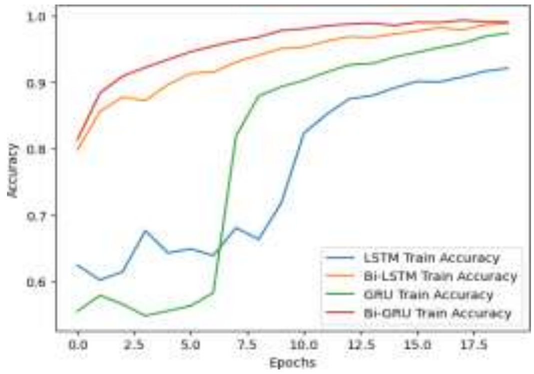
Table.5. Dataset 2

Models	Accuracy	Precision	Recall	F1 Score
LSTM	92.4	91.6	93.3	92.4
Bi-LSTM	92.9	91.2	94.9	93.0
GRU	93.1	90.8	96.0	93.3
Bi-GRU	93.1	91.1	95.5	93.3
BERT	97.2	97.8	97.8	97.8

With an accuracy of 97.5%, BERT continued to dominate Dataset 2, as indicated in Table.4, followed by Bi-GRU and GRU at 93.1%. All models performed slightly worse than Dataset 1 because to the increased variability and noise supplied by the larger dataset. However, BERT was able to effectively adjust to the diversity of the dataset because to its pre-trained embedding's. Bi-LSTM and Bi-GRU, two bidirectional models, consistently outperformed their unidirectional counterparts, proving the benefit of bidirectionality in capturing small linguistic differences.

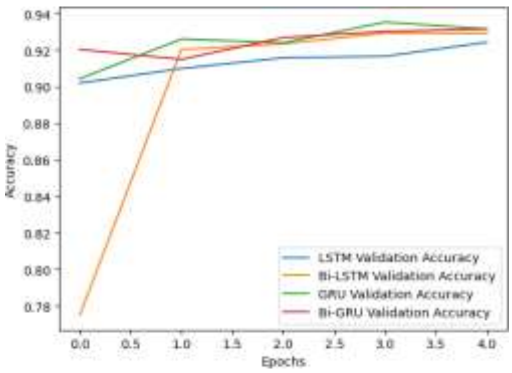


(a) Validation Accuracy

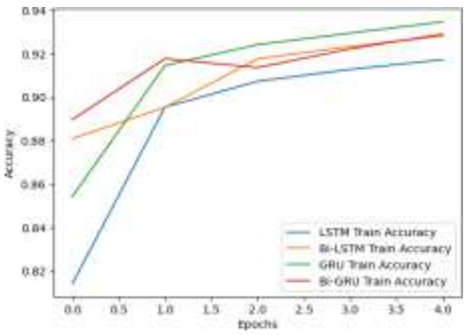


(b) Training Accuracy

Fig.4. All of the pre-trained model's accuracy and loss curves for dataset 1



(a) Validation Accuracy



(b) Training Accuracy

Fig.5. Pre-trained models accuracy and loss curves for dataset 2

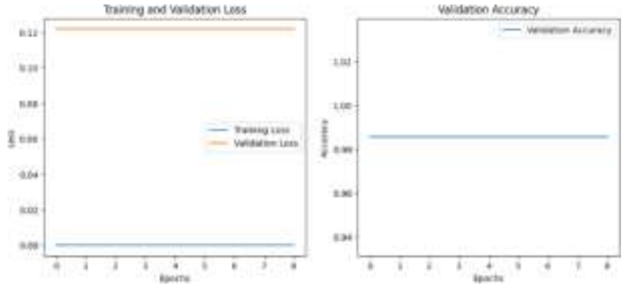


Fig.6(a) BERT model accuracy and loss curves for dataset 1

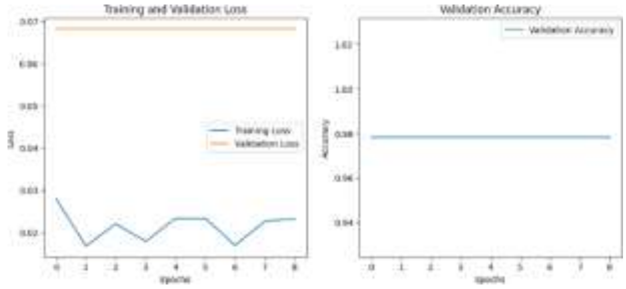


Fig.6(b) BERT model accuracy and loss curves for dataset 2

The Fig.6(a) and Fig.6(b) shows the effective optimization and generalization were demonstrated by Dataset 1, where the BERT model's accuracy curve plateaued at 98.5% and its loss curve gradually declined across training epochs with little difference between training and validation trends. Similar patterns were seen in Dataset 2, where BERT, in spite of the dataset's greater complexity, achieved convergence at 97.5% accuracy. The training and validation loss curves for BERT showed strong learning with- out appreciable over fitting in both datasets, demonstrating its versatility across a range of datasets.

6. CONCLUSION AND FUTURE WORKS

The goal of this research is to identify Reddit postings that are suicidal and those that are not. Social media is used by individuals to express their opinions, feelings, knowledge, stress, and support, especially young people who are contemplating suicide. Deep learning methods for identifying suicidal posts in social media content, especially transformer models like BERT. Suicidal and non-suicidal posts are displayed in the results. We have split article using keyword-based search since our work assumes that posts from suicide watch, anxiety, and depression are suicidal content, while posts from books, movies, popular culture, and jokes are not. However, this isn't always the case. This study categorizes English-language posts that are suicidal or non-suicidal. More supported languages could be added in the future, and text and images could be combined to enhance detection capabilities.

REFERENCES

[1] World Health Organization, "Suicide", Available at https://www.who.int/news-room/fact-sheets/detail/suicide, Accessed in 2021.

[2] PRAW Documentation, "PRAW: The Python Reddit API Wrapper", Available at

https://praw.readthedocs.io/en/stable/index.html, Accessed in 2025.

[3] Reynolds Kelly, April Kontostathis and Edwards Lynne, “Using Machine Learning to Detect Cyberbullying”, *Proceedings of International Conference on Machine Learning and Applications*, pp. 1-6, 2011.

[4] F. Haque, R.U. Nur, S.A. Jahan, Z. Mahmud and F.M. Shah, “A Transformer based Approach to Detect Suicidal Ideation using Pre-Trained Language Models”, *Proceedings of International Conference on Computer and Information Technology*, pp. 1-5, 2020.

[5] Ananthakrishnan Gayathri, Jayaraman Ashok Kumar, Trueman Tina, Mitra Satanik, A.K. Abinesh and Murugappan Abirami, “Suicidal Intention Detection in Tweets using BERT-based Transformers”, *Proceedings of International Conference on Computing, Communication and Intelligent Systems*, pp. 322-327, 2022.

[6] Rabani Syed, Khanday, Akib Khanday Ud, Khan Qamar Rayees, Hajam Ayoub, Imran Ali and Kastrati Zenun, “Detecting Suicidality on Social Media: Machine Learning at Rescue”, *Egyptian Informatics Journal*, Vol. 24, No. 2, pp. 291-302, 2023.

[7] Ji Shaoxiong, Yu Celina, Fung Sai-fu, Pan Shirui and Long Guodong, “Supervised Learning for Suicidal Ideation Detection in Online User Content”, *Complexity*, pp. 1-10, 2018.

[8] M. Chatterjee, P. Samanta, P. Kumar and D. Sarkar, “Suicide Ideation Detection using Multiple Feature Analysis from Twitter Data”, *IEEE Delhi Section Conference*, pp. 1-6, 2022.

[9] Imam Rafi, Jyoti Oishi, Afrin Zakia, Hossain Md Munawar and Mou Tamanna, “Suicidal Thought Detection using NLP on Reddit Data”, *Proceedings of International Conference on Computer and Information Technology*, pp. 1-6, 2023.

[10] A. Malhotra and R. Jindal, “Deep Learning Techniques for Suicide and Depression Detection from Online Social Media: A Scoping Review”, *Applied Soft Computing*, Vol. 130, pp. 1-7, 2022.

[11] A. Chadha, A. Gupta and Y. Kumar, “Suicidal Ideation Detection on Social Media: A Machine Learning Approach”, *Proceedings of International Conference on Technological Advancements in Computational Sciences*, pp. 685-688, 2022.

[12] Kipkebut Andrew and Emmanuel Chesire, “Deep Learning Suicide Ideation Detection Model”, *Iconic Research and Engineering Journals*, Vol. 7, No. 3, pp. 44-51, 2023.

[13] J.P. Yadav, D. Kumar and D. Chauhan, “Cyberbullying Detection using Pre-Trained BERT Model”, *Proceedings of International Conference on Electronics and Sustainable Communication Systems*, pp. 1096-1100, 2020.

[14] Mishra Ajey, Srivastava Akanksha, Modi Aashi and Pandey Ashish, “Cyberbullying and Suicide Ideation Detection via Hybrid Machine Learning Model”, *Proceedings of International Conference on Signal Processing and Integrated Networks*, pp. 265-270, 2023.

[15] K. Nikhileswar, D. Vishal, L. Sphoorthi and S. Fathimabi, “Suicide Ideation Detection in Social Media Forums”, *Proceedings of International Conference on Smart Electronics and Communication*, pp. 1741-1747, 2021.

[16] Darade Vrushali, “Deep Learning and Natural Language Processing for Suicidal Ideation using Instagram Posts”, Master of Science in Research Project, Data Analytics, National College of Ireland, pp. 1-21, 2023.

[17] O’Dea Bridianne, Wan Stephen, Batterham Philip, Calear Alison, Paris Cécile and Christensen Helen, “Detecting Suicidality on Twitter”, *Internet Interventions*, Vol. 2, No. 2, pp. 183-188, 2015.

[18] Syed Sobia and Amin Imran, “Prediction of Suicide Causes in India using Machine Learning”, *Journal of Independent Studies and Research-Computing*, Vol. 15. pp. 1-6, 2017.

[19] M.G. Harsha and A. Kaur, “A Review of Machine Learning Approaches to Assess Psychological Health: A Comparative Analysis of Non-Meditators and Meditators”, *Proceedings of International Conference on Computing Communication and Networking Technologies*, pp. 1-7, 2023.

[20] T. Bhardwaj, P. Gupta, A. Goyal, A. Nagpal and V. Jha, “A Review on Suicidal Ideation Detection based on Machine Learning and Deep Learning Techniques”, *IEEE World AI IoT Congress*, pp. 027-031, 2022.