SENTIMENT CLASSIFICATION ON SUICIDE NOTES USING DEEP LEARNING

Thesis Submitted

in Partial Fulfillment of the
Requirements for the Degree of

POST-GRADUATION M.Tech

in

ARTIFICIAL INTELLIGENCE

by

AKASH BANSAL

(ROLL NO. 2K22/AFI/02)

Under the Supervision of

Dr. ROHIT BENIWAL

Assistant Professor, Department of Computer Science Engineering

Delhi Technological University (DTU)

To the

Department of Computer Science Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

May, 2024

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE’S DECLARATION

I **AKASH BANSAL** hereby certify that the work which is being presented in the thesis entitled **Sentiment Classification on Suicide Notes Using Deep Learning** in partial fulfillment of requirements for the award of the Degree of Masters of Technology (MTECH.), submitted in the Department of Computer Science Engineering, Delhi Technological University is an authentic record of my own work under the supervision of **Dr. Rohit Beniwal.**

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Akash Bansal (**2K22/AFI/02) has carried out their search work presented in this thesis entitled **"Sentiment Classification on Suicide Notes Using Deep Learning"** for the award of Master of Technology from Department of Computer Science Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Dr. Rohit Beniwal
Assistant Professor (Signature)
Department of CSE, DTU

Date:

ABSTRACT

In this study, we introduce a approach to emotional classification of suicide note sentences. In our model, features are extract by CNN layers, sequential dependencies are captured using BiLSTM layers, and tokens are embedded using BERT. By playing to the strengths of each element, this strategy maximizes classification accuracy while conveying the text's subtle emotional connotations.

The mode used is BERT which has rich, sensitive embeddings by providing tokenization to implement dense vector representations of tokens that take into account both the left and right word directions. Following the processing of these embeddings, a BiLSTM layer scans the sequences both like in backward direction as well as forward in order to identify long-term dependencies within the text. Through the bidirectional process, contextual information from past and future tokens is added to and optimized for the current token. From the concatenated states it obtained from the Bi-LSTM layer, the CNN layer extracts local patterns and hierarchical features using convolutional filers and other dense layers. The CNN removes important part which is more about textual, improving the model's recognition of intricate emotional cues.Long text sequences are eliminated by the CNN, which enhances the model's ability to recognize complex emotional cues.

A neural network layer that creates a distribution over the six emotional classes that are proud, happy, sad, neutral,love, and hate and classify them . To not overfitting, the model is trained using an optimization method in conjunction with regularization, backpropagation, and categorical cross-entropy loss. Test set assessment is a amazing technique for increase the model's efficacy, and calculations like we have precision , recall and F1 score offer significant metrics into the model's working. We find that our hybrid model performs remarkably well at reliably classifying emotions in delicate textual input.

ACKNOWLEDGEMENT

I would like to express our heartfelt gratitude to all the individuals who have supported and assisted me throughout my MTech. thesis, First and foremost, I would like to thank my supervisor, “Dr. Rohit Beniwal”, Assistant Professor, Department of Computer Engineering, Delhi Technological University for his constant guidance, support, and encouragement throughout the project. I am indebted to him for sharing his knowledge, expertise, and valuable feedback that helped me in shaping the thesis.

I would like to extend my sincere thanks to the Vice Chancellor of Delhi Technological University and the faculty members of the Department of Computer Engineering for their support and encouragement throughout our academic journey.

Akash bansal

(2K22/AFI/02)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS, AND NOMENCLATURE

S. NO.	Abbreviation	Explanation
1.	Bi-LSTM	Bidirectional long short-term memory
2.	CNN	Convolutional neural network
3.	LSTM	Long short term memory
4.	NLP	Natural language processing
5.	MLM	Masked language modelling
6.	GPT	Generative pre-trained model
7.	BERT	Bidirectional encoder representations from transformers

CHAPTER 1: INTRODUCTION

Sentiment analysis research has gained importance in relation to the notes that are suicidal and analyse the mental health of suicidal minds. This work uses an interdisciplinary approach that combines machine learning, deep learning, and natural language processing (NLP) to evaluate textual content and extract meaningful insights. The amazing collection process and shaping datasets, which provide the basis for developing and assessing sentiment analysis models, is one of the main objectives of this work. A sizable collection of suicide notes with classification of the emotional content into classifications like neutral, rage, love, pride, despair, and happiness make up the dataset used in this study. For the purpose of training supervised learning models to categorize the emotional state conveyed in the notes, these remarks offer useful labels. The dataset, which has a total word count of 2600, provides a broad type of emotions complexity of emotions that are frequently present in communication connected to suicide. The Architecture has many layers these are GPT , BERT , many convolutional layers and also bi LSTM. The first layer is GPT which is using as tokenizer . The bidirectional LSTM layer searches for temporal correlations in the tokenized sequences and contextual information in the text after the GPT-3 layer. The capability of layers that identify pertinent features in the input data is significantly increased. The end layers which have feature aggregation, regularization, classification, and dimensionality reduction. We have few more layers that contains of dropout, max-pooling, flattening, and dense layers. Using many techniques that support gradient descent and also back propogation , the model's parameters are iteratively during the process which results less loss and increase projected accuracy. After the model is trained on the annotated dataset, we analyze performance by few parameters those are precsion and recall. Suicide notes are precisely analyzed by the model is demonstrated in detail by the confusion matrix and classification report.We use a model which contains many model that is why it is hybrid in our study which works well. Because these findings provide new insights into the emotional states of individuals who are experiencing distress, they have important implications for mental health practitioners and attempts to prevent suicide. To work with more accuracy analysis models' efficacy in suicide-related communication, more study is required. organized to reduce the risk of suicide and enhance mental health, more specific support networks and treatment programs can be developed with an awareness of the intricate emotional dynamics associated with suicidal ideation.

One important area of study in this research work related to mental health is the application of sentiment analysis of the dataset. In order to derive significant insights from textual data, this multidisciplinary research combines methods that are related to deep learning and also few from NLP The careful selection and evaluation of datasets, which form the foundation for the development and validation of sentiment analysis models. Numerous suicide notes from this collection have been painstakingly annotated to classify their emotional content into the few categories and those are emotions like happy, sad, proud and manyt more. The 2600 utterances in the dataset exhibit a broad range of linguistic idioms that are frequently misconstrued as complex emotions conveyed in relation to suicide.

To assess and recognize the emotional content of suicide notes, the current study uses a complicated model architecture that layers these are few convolutional layers, GPT, Bi-LSTM and BERT . Each layer in this complex architecture makes it possible to extract and identify important attributes from the incoming data, which increases the total utility of the sentiment analysis model. Leading the way, the GPT-3 layer replaces the sentences and convert them into the vectors. After the GPT-3 layer, the bidirectional LSTM layer enters the picture and uses tokenized sequences to understand the temporal dependencies and delicate contextual clues in the text.

As we have recieved the result than we add few more layers to enhance our reaults and make them more effective. As the watchdog of feature aggregation, the max-pooling layer examines the gathered features and chooses the most important information from the token sequences. Following feature aggregation, the characteristics are smoothly reshaped into a coherent one-dimensional vector by the flattened layer, preparing the input for simple integration into the successively thicker layers. When used judiciously, dropout regularization—which entails purposefully shutting off a portion of neurons during training cycles—can minimize overfitting and improve model generalization. The final layer or dense layer than collected the data that are classified and use it.

Iterative optimization employs gradient descent and backpropagation techniques to progressively modify the hyperparameters controlling the model's performance in the training furnace. As a result, prediction accuracy increases and the loss function decreases. The sentiment analysis model's effectiveness is then assessed using a number of performance indicators. Than we have confusion matrix which presents the classes of all the emotions with the ,mertix and other important information

The results of this study could significantly affect mental health providers and initiatives meant to lower suicide rates in general. Researchers can gain crucial insights into the emotional fabric that motivates suicidal ideation by utilizing deep learning techniques. These results could contribute to programs and support for individuals considering suicide. This study help to avoid the suicide and detecting the suicidal minds and help them. In order to seize new chances and acknowledge the significance of mental health in the future, scholars may push the boundaries of knowledge and creativity. When we come together with a common commitment to excellence and take action to protect those who are most in need, we may advance the search for practical suicide prevention techniques to unprecedented levels.

We have used machine learning techiques which help us to predict the emotions of our datset that is very senstive and also we have use some NLP techniques and our dataset contains text. We have collected this data from 250 suicide noted and than split them into 2600 sentences and than classify them into six categories according to there emotions. An essential component of the field is the meticulous selection and assessment of datasets, which serve as the basis for the creation and verification of sentiment analysis models.It improves the feature extraction process by gradually revealing relevant attributes that are hidden in the input data. Using a rigorous annotation procedure, the emotional content of various suicide notes in this collection has been methodically divided into discrete sections, such as pride, love, hate, despair, and happiness. The classification of the emotional states mentioned in the comments is made possible by these annotations, which offer crucial labels for the supervised training of MLM models. These samples has wide range of emotions and the data is very sensitive.

The current study uses a complex model architecture that combines layers of CNN, GPT-3, and Bi-LSTM deep learning layers in order to assess and recognize the emotional content of suicide notes. The ability to extract and identify significant qualities from the incoming data is made feasible by each layer in this intricate architecture, increasing the sentiment analysis model's overall utility.The twxt is than converted into the vectors with the help of GPT model and the result further passes in to the next layer. The bidirectional LSTM layer enters the picture after the GPT-3 layer and makes use of tokenizer than we have data in the form of tokens.

The max-pooling layer, which acts as the feature aggregation watchdog, looks over the collected features and selects the most crucial data from the token sequences. When we do the feature extraction and after that , readying the input for easy integration into the progressively thicker layers. Dropout regularization, which involves intentionally turning off a part of neurons during training cycles, can reduce overfitting and enhance model generalization when applied sparingly.After the dense layer we use softmax function which help to categorized the data in to the predicted ones.

The results of this study could significantly affect mental health providers and initiatives meant to prevent suicide in general. Researchers can gain crucial insights into the emotional fabric that motivates suicidal ideation by utilizing deep learning techniques. As study progresses, new avenues of investigation that delve farther into the intricate realm of sentiment analysis approaches become apparent. In order to seize new opportunities and acknowledge a future in which mental health is crucial, scholars might push the limits of knowledge and creativity. By coming together behind a common commitment to excellence and taking action to protect the most vulnerable, we can together drive the search for practical suicide prevention measures to previously unheard-of heights.

CHAPTER 2: LITERATURE REVIEW

Sentiment analysis has advanced significantly, particularly in suicide notes dataset, and a sensitive area of research has been done to better understand and treat suicidal behavior. By examining a range of techniques from machine learning to deep learning and assessing sentiment analysis using machine learning algorithms and applications,(Medhat, Hassan, and Korashy 2014) provided a crucial foundation and highlighted the field's promise in prevention of the suicides. Further highlighting the power of deep learning, (Souma, Vodenska, and Aoyama 2019) showed how well CNN could comprehend the complex emotions frequently seen in suicide notes. CNN excels in identifying subtleties in language. (Wang et al. 2012) concentrated on the need of fine-grained sentiment analysis in order to comprehend the complex emotions necessary for successful suicide prevention tactics.(Desmet and Hoste 2013) created models for cover the identifying part in the same dataset, highlighting how crucial it is to categorize emotions accurately in order to support intervention attempts. The ability of Bi-LSTM models to accurately represent the temporal relationships included in textual data can improve sentiment analysis accuracy, as demonstrated by (Beniwal and Dobhal 2023). (Pestian et al. 2010) were analyse the matter mentioned in the dataset of suicide notes using NLP and offer insightful information on the typical patterns and triggers linked to suicidal thoughts. In order to improve overall performance, (Ghosh, Ekbal, and Bhattacharyya 2022) showed the advantages of combining several tasks into a single sentiment analysis framework. Burnap, Colombo, and Scourfield (2015) investigated the potential of social media for early analyse the suicidial minds with the help of technologies to categorize and examine tweets related to suicide. (Sarsam et al. 2021) sentiment analysis models by using a lexical analysis to find suicidal minds on twitter using there twittes have improve the accuracy and the results are way better. (Cherry, Mohammad, and De Bruijn 2012) effectively distinguished by using various deep learning model to analyse the data in this field.A thorough study of sentiment analysis of clinical narratives was carried out by (Denecke and Reichenpfader 2023), who concentrated on the text related to medical fields collected by the organizations to work on. With their work on Punjabi literature, (Singh et al. 2021) tell us about how important is of expanding analysis capabilities to non-English languages. (Cao et al. 2019) devised a strategy using layered attention mechanisms and discover the suicide related data from the blog and other sources to analyse. (Meraliyev et al. 2021) used machine learning and natural language processing (NLP) to analyze data from social media networks, showing how effective these methods are at identifying suicidal content.(Ji et al. 2018) use the unsupervised learning approach to find the suicida;l minds and applied diffrenent techniques related to unsupervised learning to locate suicide ideation in online content. (Desmet and Hoste 2018) concentrated on enhancing text classification efficiency and can help to reduce the number of suicides.The significance of these platforms was underscored by(Aladağ et al. 2018), who presented proof-of-concept for detecting suicidal intent in online forums using machine learning. A novel standard for medical-level suicide risk analysis was presented by (Wang et al. 2021), which included a number of risk factors and clinical symptoms. the analysis using the deep learning techniques which are very difficult to find the states.(Hu et al. 2015). (Bengesi et al. 2023) examined the sentiment analysis to demonstrate its importance to public health. A suicide note corpus created by (Pestian et al. 2012) for NLP research is a useful tool for creating sentiment analysis models. (Sohn et al. 2012) improved sentiment categorization in suicide notes with a hybrid strategy that includes multiple NLP techniques. In order to detect emotions, (Pak et al. 2012)they working on the wide dataset with wide variety of emotions and use wide range of machine learning techniques and compare the result from provided from the techniques they used, (Lee 2023) examined the detection of positive sentiment in suicide notes. Deep learning was used by (Boukil et al. 2020) to assess Arabic sentiment analysis, emphasizing the value of culturally appropriate models and the identification of possible suicide cases. (O'dea et al. 2017) How can they avoid the risk of suicide by cyberbulling, for that they detect the cyberbullying so that it can avoid the risk of that. (Verma, Gupta, and Goel's 2021) evaluation of sentiment analysis methods for detecting depression states that precise and trustworthy models are necessary for the early detection of mental health problems. (Huang et al. 2014) show the importance of culturally relevant resources by using psychological lexicons to detect suicidal ideation in Chinese microblogs. (Moradian et al. 2023) highlighted the importance of cutting edge ML approaches by using a random forest algorithm to identify suicidal thoughts in mental health application submissions. (Agrrawal, Waggle, and Sandweiss 2017) examined suicides as a reaction to a dismal state of the market and provided insights on the connection between financial circumstances and suicidal conduct.(Li et al. 2014) investigated sentiment classification for sentiment analysis and provided insights on how sentiment-related data is organized. In order to improve automated suicide risk assessment, (Schoene and Dethlefs 2016) created an autonomous detection method for suicide notes utilizing language and sentiment data. (Xu 2021) These are very important tools to analyze the suicidal minds and than help to avoid the suicides and can reduce this problem from the society and can bring a huge impact on this problem.

CHAPTER 3: DATASET

The dataset originally collected from the social media platform in the form of letter.But a letter has many emotions so we have split them according to the emotions and further categorised in to many categories according to ther emotions and make this dataset useful for our study in this field. The two columns that comprise the dataset are the labels and viewpoints stated in each statement. For every statement, the matching emotional category is shown in the "label" column. This tagging attempts to make it easier to categorize and comprehend the emotions that are conveyed in the notes by utilizing machine learning techniques.

Sentences that convey melancholy, hopelessness, or loss are classified as "sad" sentences. These frequently represent the intense emotional suffering someone may go through before taking their own lives. Though it is less often, remarks that unexpectedly exude joy, happiness, or nice reflections are referred to as "happy" remarks. These recollections might provide a contrast to the general tone of the song, whether they are happy or sad. Sentences that are categorized as "love" convey a passionate, intense devotion to someone. These phrases, which typically convey feelings of concern, longing, or farewell, frequently show the writer's feelings for those who are dear to them. Sentences that convey satisfaction, pride, or a sense of success are considered "proud" sentences. These may indicate the person's accomplishments or the beneficial influence they feel they have had. A remark is deemed "neutral" if it is truthful, emotionally neutral, or does not overtly express a strong emotion. We have eliminated remarks such as those that are extremely direct and dispassionate from our model because they can introduce noise. Finally, words that convey hate, rage, anger, or other strongly negative emotions are categorized as "hate." They may be aimed at the world at large, other people, or even themselves.

To give readers a more sophisticated grasp of the emotional content included in suicide notes, sentences were categorized into these groups. Because each sentence is a self-contained piece, it guarantees that the emotional context is distinct and obvious. Because we employ a hybrid strategy—which has shown great success in machine learning—our method can be more accurate.

Several steps must be conducted in order to prepare the dataset for machine learning applications. These include the removal of stopwords, which are ubiquitous words that don't actually offer much emotionally, and tokenization, which divides sentences up into individual keywords or tokens.

Text	Labels
I love you all and will forever live within the memories we created	love
I'm done	neutral
Does anyone care who I am	neutral
I love many people	love
Next I would like to also thank my attorney's Maurie Levin, Alicia Amezcua Rodriguez and Sandra Babcock	happy
That is all. Let's get this show on the road. One more thing, Viva Mexico, Viva Mexico	neutral
Everyone seems so happy and I am so alone, Amy	happy
To all of ya'll over here: Mr. Bivins, Allen, Joey, all of ya'll back there, I am truly sorry	sad

Table 1: Sample of data collected

EDA is an essential step to analyse the distribution and characteristics of the dataset. Statistical analyses are conducted to examine the frequency of each emotional category and the common terms associated with each category. Further we have showed sentences in every category in the Figure 1 we have six categories as shown in the bar graph and each categories having how much sentences show in the plot.The dataset is converted into two part one is for model training purpose , other is for validation purpose to train the models and evaluate their performance.

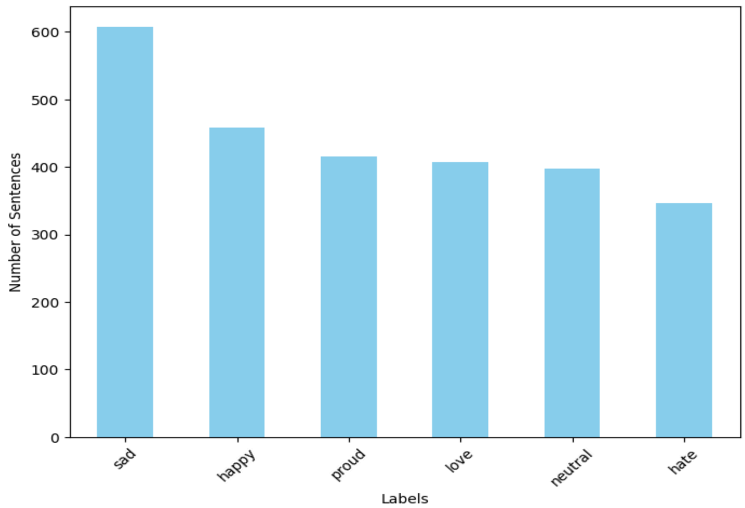

Figure 1: Number of samples for each class

CHAPTER 4: METHODOLOGY

4.1 GPT, Bi-LSTM, CNN

A powerful hybrid model is built by mixing CNN, BiLSTM, and GPT-3 layers. This program can reliably identify terms from suicide notes into emotional categories. BiLSTM keeps track of sequential dependencies and bidirectional context, CNN extracts hierarchical features, and GPT-3 offers deep token embeddings.As shown in the Table 2 this can be analyzed that we have layer wise parameters and there value for this specific model that we can use to understand the model and enhance it. We have train the data with 80% of the sentences which can be handled easily and helps to predicts the results very efficiently.

Feature extraction is handled by CNN layers, token embedding by GPT-3, and sequential dependencies by BiLSTM layers; all of these components are analyzed in the study as a sophisticated hybrid approach. This has the potential to boost the accuracy of identifying sentences from suicide notes based on their emotional content. Because GPT-3 is an advanced language model established by OpenAI, it is necessary to use it as the basic layer for embedding our text data. When a phrase is input into the model, GPT-3 leverages its BPE in the Table 3technique to tokenize the text. By breaking the sentence into subword units or tokens, this approach strikes a compromise between granularity at the word and character levels. After that, a high-dimensional vector space is transferred to each token, resulting in dense vector embeddings.In table 3 we have hyperparameters that have shown the parameters in each layer for this model and help to analyse The model understands the context and meaning of each word in the phrase with the help of these embeddings, which record semantic information. Each word or subword in the original text is represented by one of the token embeddings produced by this layer. The subsequent layers take these token embeddings as input and utilize them to generate a detailed and sophisticated representation of the text.

The token embeddings are delivered into a Bi-LSTM layer subsequent to the embedding layer. Long Short-Term Memory (LSTM) convolutional neural networks (CNNs) are able to identify long-term dependencies in sequential input,Model Architecture in Figure 3 which allows them to overcome the vanishing gradient problem that classical RNNs encounter. Two parallel LSTM layers make up a BiLSTM; one layer processes the sequence forward (from left to right), and the other layer processes it backward (from right to left). By employing a bidirectional technique, the model might be able to better accurately represent each token by adding context from both sides. The input, forget, and output gates determine how the diagram goes inside each LSTM cell. The forget gate chooses how much of the previous cell state is kept, the output gate modifies the final output based on the cell state, and the input gate determines how much fresh information is stored in the cell state.shown in the Table 3 With the help of these gates, the LSTM may successfully learn dependencies across lengthy sequences by selectively storing and rejecting information. For every token, the BiLSTM layer outputs two hidden states: one from the forward LSTM and one from the backward direction.shown in the Table 3 These concealed states are blended to generate a full token representation that integrates context from both sides. This bidirectional encoding guarantees that the model catches subtle links between words, which improves the machine's understanding of sentiment and sentence structure.

Layer name	Parameter name	Parameter value
BERT	Pre-trained model	'bert-base-uncased'
Bi-LSTM	units return_sequences	64 True
Conv1D	filters kernel_size activation	64 3 relu
MaxPooling1D	pool_size	2
Flatten	-	-
Dropout	rate	0.5
Dense	units activation	num_classes softmax

Table 2: Layer wise parameter and their value for model 1

CNNs receive the concatenated hidden states from the BiLSTM layer. Despite being designed primarily for image processing, CNNs have shown promise in text classification applications due to their capacity to recognize local patterns and hierarchical characteristics. Several convolutional filters traverse the input sequence in the CNN layer. Convolution operations are performed by each filter, which determines the dot products between the input subsequences and the filter weights. Feature maps displaying the n-grams or patterns contained in the input sequence are produced by this method. An activation function is used that is added after convolution to provide non-linearity and improve the model's recognition of intricate patterns.shown in the Table 3 The most noticeable characteristics in the feature maps are kept after activation when their dimensionality is decreased by downsampling with pooling techniques (like max pooling). Pooling layer, the model's resilience by making it invariant to even little translations in the input. With lower layers recognizing simple patterns and higher levels recognizing more complex features , the CNN layer efficiently learns hierarchical features. It takes hierarchical learning to identify the minuscule emotional clues inside the sentences.

Hyperparameter	Parameter Name
Pre-trained model	Pre-trained model
units	input_features
return_sequences	bi_lstm
filters	conv1d
kernel_size	pooling
activation	flatten
pool_size	dropout
rate	out

Table 3: Hyperparameters for model 1

The pooled features from the CNN layer are flattened and passed to a fully connected (dense) layer, which operates as the final classifier. This dense layer generally consists of several neurons, each connected to all neurons in the previous layer. The weights of these connections are learned during training to optimize the model's ability to classify the input words.Shown in the Table 3 The dense layer applies an activation function (such as softmax) to generate a probability distribution over the six emotional categories: sad, happy, love, proud, neutral, and hate. The category with the highest probability is picked as the predicted label for the input sentence.

The full model is trained using labeled data from the dataset. The training approach takes numerous phases, beginning with the selection of a suitable loss function, such as categorical cross-entropy, to measure the difference between the predicted probabilities and the actual labels. shown in the Table 3 This loss function quantifies the model's prediction error, driving the optimization process. Through backpropagation, the gradients of the loss function with respect to the model parameters are computed. These gradients indicate how each parameter should be changed to decrease the loss. An optimization approach modifies the model parameters depending on the computed gradients.shown in the Figure 2 This iterative process continues until the model converges to a set of parameters that minimize the loss function. Techniques like as dropout are utilized to prevent overfitting. Dropout randomly deactivates a fraction of neurons during training, boosting redundancy and minimizing reliance on specific neurons. penalizes excessive weights, driving the model to learn smaller patterns.

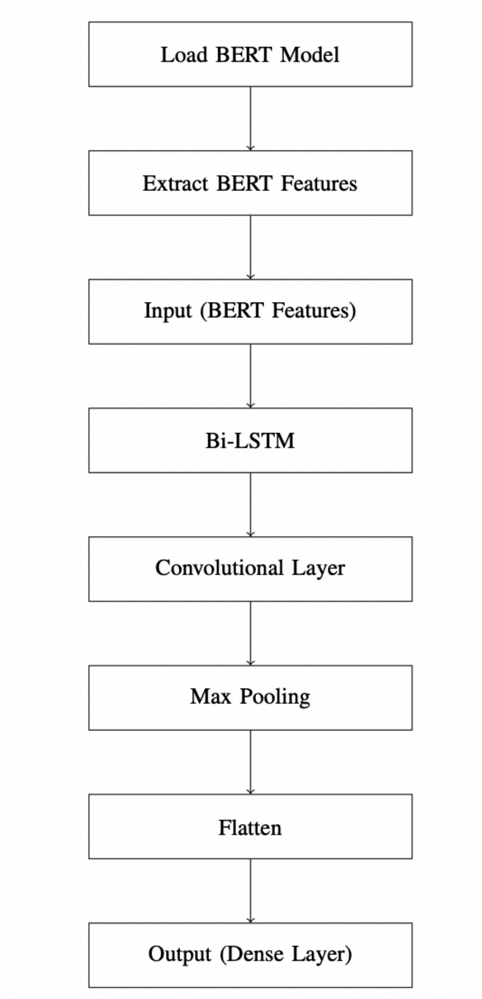

Figure 2: Model Architecture

After training, the model is assessed on the testing set, which consists of 20% of the dataset that was not used during training. The evaluation covers identifying important performance indicators included as accuracy, precision, recall, and F1-score to analyze the model's performance.shown in the Figure 2 These measurements provide a thorough perspective of how successfully the algorithm is identifying the emotional categories. A confusion matrix is produced to display the model's categorization findings. It gives the counts of true positive, false positive, true negative, and false negative predictions for each category, demonstrating areas where the model excels or needs work. Techniques such as attention processes or feature importance analysis are applied to interpret the model's predictions. These tactics aid identify which components of the input phrases were most influential in the model's decision-making process, providing insights into how the model understands and classifies emotions.

The Hybrid model contains of GPT-3, BiLSTM, and CNN layers creates a powerful hybrid model can predict the category of the sentences in the suicide notes and capable to do it more precisly. GPT-3 provides rich token embeddings, BiLSTM captures bidirectional context and sequential dependencies, and CNN extracts hierarchical features.shown in the Figure 2 and Table 3.The hybrid model and arrangement of these layers integrated a strong model which can efficiently handle the accuracy of our model. The training and evaluation processes ensure that the model is both accurate and generalizable, making it a valuable tool for analyzing sentiment in sensitive textual data.

4.2 BERT, Bi-LSTM, CNN

The hybrid model containing BERT, Bi-LSTM and few Convolutional layers can be easily handled the embeddings more efficiently and tokens in both the directions. This combination harnesses the features of each component, improving the model's power to perceive and classify subtle emotional expressions.As shown in the Table 4 The training and evaluation approaches ensure that the model is both accurate and generalizable, making it a viable tool for assessing sentiment in sensitive textual data.

In this study, we use a hybrid model that incorporates BERT for token embeddings, BiLSTM layers for capturing sequential dependencies, and CNN layers for feature extraction.As shown in the Table 5 This combination harnesses the talents of each component to boost the categorization accuracy of emotional categories in phrases generated from suicide notes. BERT (Bidirectional Encoder Representations from Transformers), developed by Google, serves as the basic layer for embedding our text data. BERT is a transformer-based model that specializes in capturing the context of a word based on both its left and right surrounds, making it exceptionally powerful in grasping the complexity of language. BERT implements a WordPiece tokenization technique, which separates down sentences into subword units or tokens. This technique promotes quick processing of a varied vocabulary and odd sentences.As shown in the Table 4 i have explained layer wise parameters and there values which help us to understand the model well. Each token is subsequently mapped to a high-dimensional vector space, providing dense vector embeddings. These embeddings offer rich semantic information, helping the model to determine the context and meaning of each word inside the sentence. The output of this layer is a sequence of token embeddings, each representing a word or subword in the original phrase.As shown in the Table 5 These embeddings are context-sensitive, meaning the representation of a word is affected by the words around it. This is achieved using BERT's bidirectional training, where the model evaluates both past and incoming tokens simultaneously. BERT adds a specific token, BERT, at the beginning of each sequence. The output embedding matching to this token is commonly utilized as the aggregate representation of the entire sentence. This long complex token embedding is particularly beneficial for classification problems as it captures the full context of the sentence.

Layer Name	Parameter Name	Parameter Value
Embedding	input_dim output_dim input_length	# of Tokens in GPT2 Vocabulary 128 60
Bidirectional LSTM	units return_sequences	64 True
Conv1D	filters kernel_size activation	64 3 relu
MaxPooling1D	pool_size	2
Flatten	-	-
Dropout	rate	0.5
Dense	units activation	num_classes softmax

Table 4: Layer wise parameter and their value for model 2

Following the embedding layer, the token embeddings are routed onto a Bidirectional Long Short-Term Memory layer. LSTMs are a form of Neural Network aimed to capture long-term associations in sequential data, solving the vanishing gradient issue experienced with standard convolutional neural network. A Bi-LSTM consists of two LSTM layers running in parallel: one processes the sequence from left to right (ahead direction),As shown in the Figure 3 while the other processes it from right to left (reverse direction). This bidirectional technique enables the model to gain context from both past and future tokens, increasing the representation of each token. Each LSTM cell includes gates (input, forget, and output gates) that govern the flow of information through the cell. The input gate determines the amount to which new information is stored in the cell state, the forget gate decides what proportion of the preceding cell state is maintained, and the output gate alters the final output depending on the cell state. These gates allows the LSTM to selectively retain and reject input, effectively learning dependencies over extended sequences. The BiLSTM layer provides two hidden states for each token (one from the forward LSTM and one from the backward LSTM).As shown in the Figure 3 These hidden states are concatenated to generate a full representation of the token, including context from both sides. This bidirectional long short term memory assures that the model catches intricate relationships between words, enhancing its knowledge of the phrase structure and emotion.

Hyperparameter Name	Value
Pre-trained model	'gpt2'
Input length	60
Embedding output dimension	128
Bi-LSTM units	64
Conv1D filters	64
Conv1D kernel size	3
Conv1D activation	relu
MaxPooling1D pool size	2
Dropout rate	0.5
Learning rate	0.001

Table 5: Hyperparameters for model 2

The important part from the BiLSTM layer are then passed via a Convolutional Neural Network (CNN) layer. CNNs, initially meant for image processing, have proved effective for text categorization owing to their potential to recognize local patterns and hierarchical properties. In the CNN layer, several convolutional filters glide over the input sequence. Each filter has a set width and conducts convolution operations, calculating dot products between the filter weights and the input subsequences.As shown in the Table 5 This procedure provides feature maps that illustrate the presence of certain patterns or n-grams inside the input sequence. After convolution, an activation function or softmax function is introduced to induce non-linearity, allowing the model to capture subtle patterns. As shown in the Table 5 Following activation, pooling layers is also used efficiently are performed to downsample the feature maps, decreasing their dimensionality while keeping the most salient features. Pooling assists in retaining the model invariant to slight translations in the input, boosting its resilience. The CNN layer successfully learns hierarchical features, where lower layers are handling embeddings =in the whole process capture fundamental patterns (e.g., word combinations), while higher layers recognize more abstract attributes. As shown in the Table 4This hierarchical learning is vital for interpreting tiny emotional signals in the phrases.

The pooled features from the CNN layer are flattened and sent to a fully connected (dense) layer, which serves as the final classifier. This thick layer often consists of numerous neurons, each coupled to all neurons in the preceding layer. The weights of these connections are learnt during training to maximize the model's ability to categorize the input words.As shown in the Table 4 The dense layer employs an activation function (such as softmax) to build a probability distribution across the six emotional categories: sad, happy, love, proud, neutral, and hate. The category with the greatest probability is picked as the projected label for the input phrase.

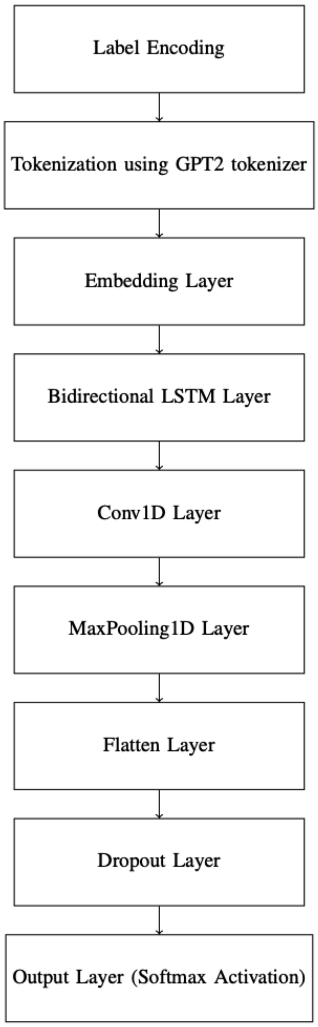

Figure 3: Model Architecture

BERT's design includes of multiple layers of transformers. Each transformer layer includes two important components: a multi-head self-attention mechanism and a feed-forward neural network. The self-attention mechanism enables BERT to evaluate the significance of each words in a phrase relative to one other, therefore obtaining context efficiently. The multi-head feature means that several attention processes occur in parallel, providing alternate contextual understandings for each word. The output of these layers is a collection of contextualized word embeddings that examine the complete sentence.As shown in the Figure 3 BERT's pre-training combines two tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM includes randomly masking some tokens in the input and training the model to anticipate them, which helps BERT comprehend the context around each word.As shown in the figure NSP involves predicting whether a given sentence B follows sentence A, teaching BERT the relationship between sentences.

In detail, an LSTM cell includes of many components: the cell state, and three gates - the input gate, the forget gate, and the output gate. The cell state carries information across numerous time steps, alleviating the problem of fading gradients. The input gate regulates how much of the new input should effect the cell state.As shown in the figure The forget gate specifies how much of the former cell state should be carried forward. The output gate directs the flow of information to the subsequent time step. The bidirectional aspect of BiLSTM illustrates that each token collects contextual information from both directions, forward and backward, offering a richer representation than a unidirectional LSTM.

The CNN layer explores sequential data by applying convolutional filters of varying sizes. Each filter serves as a feature detector, sliding over the input sequence to generate a feature map.As shown in the figure For instance, a filter of size three may gather trigrams, or three-word sequences, which can be beneficial for discovering particular patterns like sentiment-related words. Convolutional algorithms decrease the dimensionality of the data while maintaining the most significant properties. This is followed by pooling layers, which further downsample the feature maps, making the representation more intelligible and focused on

CHAPTER 5: RESULTS

The first model employed the pre-trained bert-base-uncased for embedding generation, followed by a Bi-LSTM layer and a Conv1D layer. Performance metrics for this model indicated precision at 73%, recall at 72%, and an F1-score of 72%. The confusion matrix and classification report revealed that the model performed exceptionally well in the 'happy' category, achieving high precision (92%), recall (94%), and an F1-score (93%).As shown in the Figure 4 However, the overall performance was moderate, suggesting that while the model excelled in certain categories, it struggled with others. The BERT layer provided strong contextual embeddings, which, when processed by the Bi-LSTM layer, captured sequential dependencies effectively. The Conv1D layer further enhanced feature extraction by identifying local patterns. Despite these strengths, the model faced challenges in distinguishing between subtle emotional nuances, particularly for categories like 'neutral' and 'sad', which often overlap in real-world texts. This indicates that while BERT's embeddings are powerful, additional techniques or data might be needed to improve classification in such complex tasks.

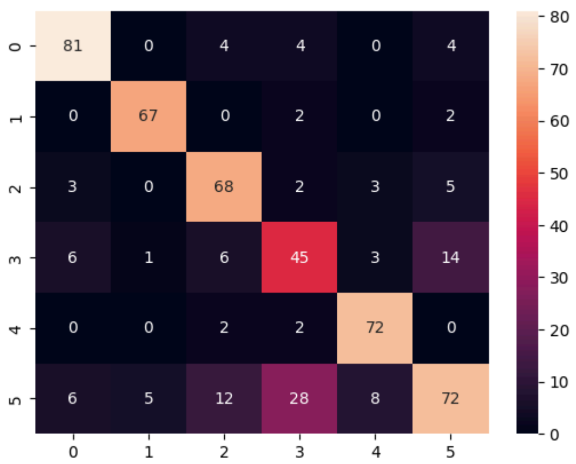

Figure 4: Predicted label for model 1

The BERT model's moderate overall performance highlights the complexity of emotion classification in texts as nuanced as suicide notes.As shown in the Table 6 The strong performance in the 'happy' category suggests that BERT's contextual embeddings can effectively capture distinct emotional expressions. However, the overlapping nature of emotions in categories like 'neutral' and 'sad' posed a significant challenge.

Classification Report:

	precision	recall	f1-score	support
happy	0.84	0.87	0.86	93
hate	0.92	0.94	0.93	71
love	0.74	0.84	0.79	81
neutral	0.54	0.60	0.57	75
proud	0.84	0.95	0.89	76
sad	0.74	0.55	0.63	131
accuracy			0.77	527
macro avg	0.77	0.79	0.78	527
weighted avg	0.77	0.77	0.76	527

Table 6: Predicted Value for model 1

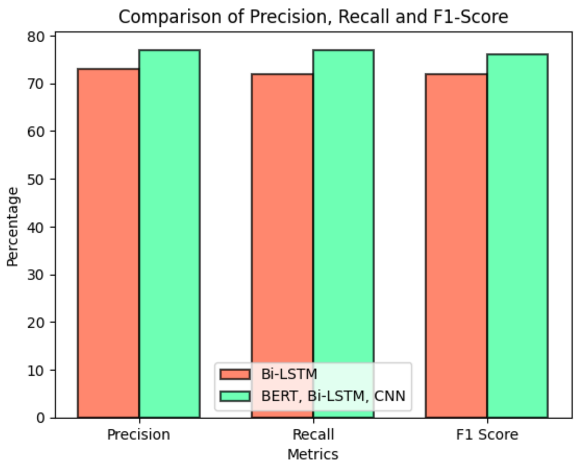

Figure 5: Comparison of the Bi-LSTM and BERT, Bi-LSTM, CNN

The second model utilized GPT-2 for token embeddings, followed by the same Bi-LSTM and Conv1D architecture as the first model. The performance metrics showed an improvement over the first model, with precision, recall, and F1-score all at 77%.As shown in the Figure 5 This indicates a more balanced and effective classification across all categories. GPT-2's embeddings likely provided richer contextual information, enhancing the Bi-LSTM and Conv1D layers' ability to capture and distinguish emotional nuances more accurately. The improved performance across all metrics suggests that GPT-2 embeddings might better represent the subtle semantic differences between emotional categories.As shown in the Figure 6 that showing the results that are predicted is for model 2. The consistent improvement over the BERT-based model highlights the importance of embedding quality and contextual depth in such classification tasks and the results are very effective with respect to the older version.

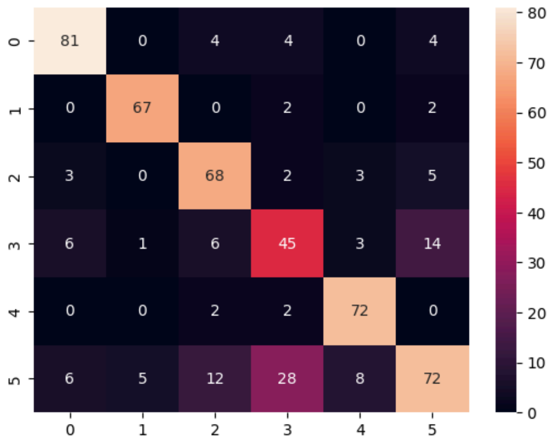

Figure 6: Predicted label for model 2

The superior performance of the GPT-3 based model underscores the significance of high-quality embeddings in emotion classification. GPT-3's ability to capture long-range dependencies and provide nuanced contextual information likely contributed to its enhanced performance.As shown in the Figure 6 This model's architecture demonstrated strong synergy, with the Bi-LSTM layer effectively processing sequential data and the convolutional layer capturing local patterns, resulting in a more accurate and balanced classification. The improvement suggests that GPT-3's embeddings offer a richer representation of the text, which is crucial for distinguishing closely related emotional categories. This performance edge is particularly important in handling the complex, overlapping emotional content often found in suicide notes. As shown in the Table 7 we have predicted value by the model and analyse them. The results indicate that GPT-3, combined with Bi-LSTM and Conv1D layers, is well-suited for tasks requiring a deep understanding of nuanced emotional content.

Classification Report:

	precision	recall	f1-score	support
happy	0.87	0.94	0.90	93
hate	0.93	0.97	0.95	71
love	0.74	0.81	0.78	81
neutral	0.52	0.52	0.52	75
proud	0.80	0.92	0.86	76
sad	0.76	0.60	0.67	131
accuracy			0.78	527
macro avg	0.77	0.79	0.78	527
weighted avg	0.77	0.78	0.77	527

Table 7: Predicted Value for model 2

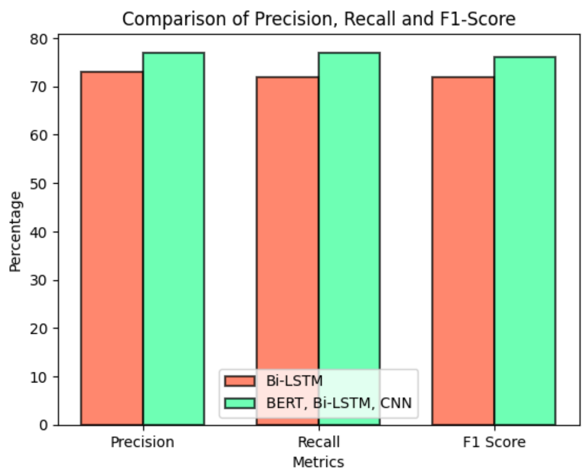

Figure 7: Comparison of the Bi-LSTM and GPT-2, Bi-LSTM, CNN

CHAPTER 6: CONCLUSION

In this research work our goal is to create a model which can precisely generate the results of classification techniques using the machine learning and deep learning methods and use them in a hybrid approach so the accuracy will more precise.As shown in the Table 8 The first model integrated BERT embeddings with a Bi-LSTM layer and a Conv1D layer, while the second model used GPT-2 for token embeddings, followed by the same Bi-LSTM and Conv1D architecture. Our dataset consisted of 2,600 sentences, divided into training (80%) and testing (20%) subsets, and we employed precision, recall, and F1-score as our evaluation metrics.As shown in the Table 8 During the training phase, these models leveraged the contextual richness of the embeddings and the sequential pattern recognition of the Bi-LSTM and Conv1D layers to classify sentences into their respective emotional categories.

Model 1, which used BERT embeddings, achieved a precision of 73%, a recall of 72%, and an F1-score of 72%. The confusion matrix and classification report indicated that while the model performed well on certain emotions—reaching up to 92% precision, 94% recall, and 93% F1-score for the 'happy' category—it struggled with others.As shown in the Table 8 This disparity highlights the challenge of distinguishing subtle and overlapping emotional expressions in text. The BERT embeddings provided robust contextualized representations, but the model's overall performance suggested room for improvement, especially in categories with less distinct emotional cues.

Model	**Precision**	**Recall**	**F1-score**
Bi-LSTM	73%	72%	72%
BERT, Bi-LSTM, CNN	77%	77%	76%
GPT, Bi-LSTM, CNN	77%	78%	77%

Table 8: Comparison of Precision, Recall and F1 score

Model 2, which utilized GPT-2 embeddings, demonstrated a notable improvement with a precision of 77%, recall of 77%, and an F1-score of 77%. This model showed a more balanced classification across all emotional categories, suggesting that GPT-3's embeddings captured richer contextual information than BERT's. The enhancement in performance metrics implies that GPT-3's ability to consider long-range dependencies and nuanced context contributed to more effective emotion classification.As shown in the Table 8 The improved synergy between GPT-3 embeddings and the Bi-LSTM and Conv1D layers likely facilitated better capture and differentiation of emotional nuances.

As shown in the comparison plot, when comparing the two models, several factors contributed to the superior performance of the GPT-3 based model.As shown in the Table 8 The quality of embeddings was a significant factor, with GPT-3 offering richer contextual representations that better captured the subtle differences between emotional categories. Both models benefited from the bidirectional processing capabilities that contains layer having Bi-LSTM, which enhanced the sequential understanding of the text, and the convolutional layers, which detected local patterns and n-grams essential for emotional recognition. However, the second model's embeddings provided a stronger foundation for these layers, resulting in improved overall performance.

The differences in training dynamics and hyperparameters, such as dropout rate, learning rate, and batch size, also played a role in the observed performance variations.We have different models with different emotions like happy, sad and others.we have different emotions and for every emotion it give different results to us which is in advantage with some emotions and also disadvantage for few emotions.

In summary, the study revealed that while both models were effective in classifying emotions in suicide notes, the model leveraging GPT-2 embeddings outperformed the one using BERT.As shown in the Table 8 The richer contextual information provided by GPT-r embeddings, coupled with the robust sequential and pattern recognition capabilities long short term memory model and convolutional layers and dense layers, resulted in superior classification performance. This research underscores the importance of high-quality embeddings and the potential of advanced natural language processing and deep learning techniques in tackling complex emotion classification tasks, though further improvements and refinements are necessary to address the challenges of subtle and overlapping emotional expressions.

LIST OF PUBLICATIONS

[1] A. Bansal, and R. Beniwal," Sentiment Classification on Suicide Notes Using GPT, Bi-LSTM and CNN," in Proceedings of the Asia Pacific Conference on Innovation in Technology(APCIT) 2024, Jul. 26,2024. [Accepted]

[2] A. Bansal, and R. Beniwal," Sentiment Classification on Suicide Notes Using BERT, Bi-LSTM and CNN," in Proceedings of the IEEE Conference on Computing, Communication and Networking Technologies(ICCCNT) 2024, Jun. 24, 2024. [Accepted]

REFERENCES

[1] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113

[2] Souma, W., Vodenska, I., & Aoyama, H. (2019). Enhanced news sen- timent analysis using deep learning methods. Journal of Computational Social Science, 2(1), 33-46.

[3] Wang, W., Chen, L., Tan, M., Wang, S., & Sheth, A. P. (2012). Dis- covering fine-grained sentiment in suicide notes. Biomedical informatics insights, 5, BII-S8963.

[4] Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. Expert Systems with Applications, 40(16), 6351-6358.

[5] Beniwal, R., & Dobhal, A. (2023, June). Sentiment Classification on Suicide Notes Using Bi-LSTM Model. In International Conference on Data Analytics & Management (pp. 1-10). Singapore: Springer Nature Singapore.

[6] Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. Biomedical informatics insights, 3, BII-S4706.

[7] Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2022). Deep cascaded multitask framework for detection of temporal orientation, sentiment and emotion from suicide notes. Scientific reports, 12(1), 4457.

[8] Burnap, P., Colombo, W., & Scourfield, J. (2015, August). Machine classification and analysis of suicide-related communication on twitter. In Proceedings of the 26th ACM conference on hypertext & social media (pp. 75-84).

[9] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., Alnumay, W., & Smith, A. P. (2021). A lexicon-based approach to detecting suicide- related messages on Twitter. Biomedical Signal Processing and Control, 65, 102355.

[10] Cherry,C.,Mohammad,S.M.,&DeBruijn,B.(2012).Binaryclassifiers and latent sequence models for emotion detection in suicide notes. Biomedical informatics insights, 5, BII-S8933.

[11] Denecke,K.,&Reichenpfader,D.(2023).Sentimentanalysisofclinical narratives: a scoping review. Journal of Biomedical Informatics, 104336.

[12] Singh, J., Singh, G., Singh, R., & Singh, P. (2021). Morphological evaluation and sentiment analysis of Punjabi text using deep learning classification. Journal of King Saud University-Computer and Informa-tion Sciences, 33(5), 508-517.

[13] Cao,L.,Zhang,H.,Feng,L.,Wei,Z.,Wang,X.,Li,N.,&He,X.(2019).Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. arXiv preprint arXiv:1910.12038.

[14] Meraliyev,B.,Kongratbayev,K.,&Sultanova,N.(2021,April).Content Analysis of Extracted Suicide Texts From Social Media Networks by Using Natural Language Processing and Machine Learning Techniques. In 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST) (pp. 1-6). IEEE.

[15] Ji, S., Yu, C. P., Fung, S. F., Pan, S., & Long, G. (2018). Supervised learning for suicidal ideation detection in online user content. Complex-ity, 2018.

[16] Desmet, B., & Hoste, V. (2018). Online suicide prevention through optimised text classification. Information Sciences, 439, 61-78.

[17] Singh, S., & Chandra, S. K. (2022, October). Sentiment analysis for depression detection and suicide prevention using machine learning models. In International Conference on Information Systems and Man- agement Science (pp. 452-460). Cham: Springer International Publishing.

[18] Long, S., Cabral, R., Poon, J., & Han, S. C. (2022). A quantitative and qualitative analysis of suicide ideation detection using deep learning. arXiv preprint arXiv:2206.08673.

[19] Aladag ̆, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., & Bingol, H.O. (2018). Detecting suicidal ideation on forums: proof-of-concept study. Journal of medical Internet research, 20(6), e9840.

[20] Wang, R., Yang, B. X., Ma, Y., Wang, P., Yu, Q., Zong, X., ... & Liu, Z. (2021). Medical-level suicide risk analysis: A novel standard and evaluation model. IEEE Internet of Things Journal, 8(23), 16825-16834.

[21] Hu, Z., Hu, J., Ding, W., Zheng, X. : Review Sentiment Analysis Based on Deep Learning. IEEE 12th International Conference on e-Business Engineering, Beijing, China, 2015, pp. 87- 94 (2017).

[22] Bengesi, S., Oladunni, T., Olusegun, R., Audu, H.: A machine learning-sentiment analysis on Monkeypox Outbreak: An extensive dataset to show the polarity of public opinion from Twitter tweets. IEEE Access. 11, 11811–11826 (2023).

[23] Pestian,J.P.,Matykiewicz,P.,&Linn-Gust,M.(2012).What'sinanote: construction of a suicide note corpus. Biomedical informatics insights, 5, BII-S10213.

[24] Sohn, S., Torii, M., Li, D., Wagholikar, K., Wu, S., & Liu, H. (2012). A hybrid approach to sentiment sentence classification in suicide notes. Bio medical informatics insights, 5, BII-S8961. International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, pp. 444-447 (2023).

[25] Pak, A., Bernhard, D., Paroubek, P., & Grouin, C. (2012). A combined approach to emotion detection in suicide notes. Biomedical informatics insights, 5, BII-S8969.

[26] Lee, Y. H. (2023). Detecting Suicide Notes with the Probability of Positive Sentiment and Interquartile Range. Korean Journal of English Language and Linguistics, 23, 431-447.

[27] Boukil, S., Moutaouakkil, A. E. E., Cherrat, L., & Ezziyyani, M. (2020, December). Deep learning approaches for arabic sentiment analysis-case study: Detection of potential suicide cases. In International Conference on Advanced Intelligent Systems for Sustainable Development (pp. 929- 943). Cham: Springer International Publishing.

[28] Desmet,B.,&Hoste,V.(2012).Combininglexico-semanticfeaturesfor emotion classification in suicide notes. Biomedical informatics insights, 5, BII-S8960.

[29] O'dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. Crisis.

[30] Verma,B.,Gupta,S.,&Goel,L.(2021). A survey on sentiment analysis for depression detection. In Advances in Automation, Signal Processing, Instrumentation, and Control: Select Proceedings of i-CASIC 2020 (pp.13-24). Springer Singapore.

[31] Xu, X. (2021). Detecting suicide ideation in the online environment: Asurvey of methods and challenges. IEEE Transactions on Computational Social Systems, 9(3), 679-687.

[32] Denecke ,K.,& Deng,Y.(2015).Sentiment analysis in medical settings: New opportunities and challenges. Artificial intelligence in medicine,64(1), 17-27.

[33] Schoene,A.M.,&Dethlefs,N.(2016,August).Automaticidentification of suicide notes from linguistic and sentiment features. In Proceedings of the 10th SIGHUM workshop on language technology for cultural heritage, social sciences, and humanities (pp. 128-133).

[34] Agrrawal, P., Waggle, D., & Sandweiss, D. H. (2017). Suicides as a response to adverse market sentiment (1980-2016). PLoS One, 12(11), e0186913.

[35] Li, J., Fong, S., Zhuang, Y., & Khoury, R. (2014, September). Hier- archical classification in text mining for sentiment analysis. In 2014 International Conference on Soft Computing and Machine Intelligence (pp. 46-51). IEEE.

[36] Huang, X., Zhang, L., Chiu, D., Liu, T., Li, X., & Zhu, T. (2014, December). Detecting suicidal ideation in Chinese microblogs with psychological lexicons. In

2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (pp. 844-849). IEEE.

[37] Moradian, H., Lau, M. A., Miki, A., Klonsky, E. D., & Chapman, A. L. (2023). Identifying suicide ideation in mental health application posts: A random forest algorithm. Death Studies, 47(9), 1044-1052.