Quest Journals
Journal of Electronics and Communication Engineering Research
Volume 8 ~ Issue 5 (2022) pp: 64-71
ISSN(Online) : 2321-5941
www.questjournals.org

Research Paper

Suicidal Thoughts Prediction from Social Media Posts using Machine Learning and Deep Learning

Siddhi Shah, Samrudhi Kadam, Subhadra Pandhare, Aishwarya Ambapkar, Dr. Smriti H. Bhandari

(Department of Computer Science and Engineering, Annasaheb Dange College of Engineering & Technology) Corresponding Author: Siddhi Shah

ABSTRACT: *Though significant advancements in the diagnosis and treatment are evident in the field of mental disorders, suicide remains a critical public health problem. Suicidal tendencies of a person can be observed through his/her behavior. Now-a-days many individuals are using social media platforms to share their thoughts and even discuss their problems. Analyzing these social media posts considering language preferences and topic descriptions may help in early detection of suicidal tendencies leading to suicide prevention. The proposed work attempts to address this issue using machine learning and deep learning methodologies for classifying textual social media posts to detect the presence of possible suicidal ideation. Natural Language Processing is used for pre-processing the raw social networking text data. For the text classification task Machine learning techniques such as Support Vector Machine (SVM), Bernoulli Naive Bayes (NB) algorithm, Decision Tree (DT) classifier, Random Forest (RF), Extreme Gradient Boosting (XG Boost), Stochastic Gradient Descent (SGD) and K-Nearest Neighbors (K-NN) algorithms are used. Further deep learning neural networks like Long Short Term Memory (LSTM), Convolutional Neural Network (CNN), Bidirectional LSTM and Attention mechanisms are explored to achieve improved performance. Experimental results show that BI-LSTM with Attention mechanism provides 93.75% overall accuracy with Twitter dataset.*
KEYWORDS: *suicide ideation; early suicide detection; twitter dataset.*

***R**eceived 12 May, 2022; **R**evised 24 May, 2022; **A**ccepted 26 May, 2022 © The author(s) 2022.*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

Nearly 800,000 people resort to suicide each year. Suicide is a common cause of death, with an overall suicide commitment rate of 10.5 per 100,000 individuals. Many people fall a prey to suicide every year making early detection and suicide prevention an important issue for the society. Recently people have been posting about their thoughts and views over various topics on social media platforms like Twitter and Reddit. The online writing community also provides a platform for users to publicly express their thoughts and opinions and share social messages. Behaviors and patterns exhibited in public online posts have been observed to have a link to the real life situations and suicidal attempts. The goal of this work is to build machine learning and deep learning models to carry out the text classification task of predicting suicidal thoughts from social media content. It consists of two approaches to textual data mining methods. The first one consists of machine learning methodology wherein data pre-processing with NLP and feature extraction is done. The resultant features are further to train traditional machine learning systems. The second framework includes various deep learning neural networks. It involves pre-processing the textual data containing social media posts, extracting of features using word embedding, followed by deep learning architectures for Long Short Term Memory, Convolutional neural networks and Attention mechanisms.

II. LITERATURE SURVEY

In the previous studies on suicide intention understanding and prevention, researchers have focused on the psychological and clinical aspects of suicide. However, recently, natural language processing methods and supervised learning methods are being studied. Some of these studies analyzed posts from suicide blogs and social media websites. However, these studies have some limitations. Michael Mesfin Tadess et al. [1] used Reddit social media data to look for signs of suicidal ideation. They demonstrated a stronger association between suicidal ideation and language use preferences by applying different methods of natural language processing and text classification.

Israel Aminu [2] built a suicidal ideation classifier which predicted whether a post is likely to be suicidal or not. For building the model, he used the stochastic gradient descent algorithm which gave an accuracy of 91%. Suicidal Ideation Detection by Shaoxiong et al. [3] focuses on text-based methods using deep learning for feature learning. A research paper published by Suyash Dabhane [4] focuses on ensemble learning method which can detect depression more accurately. They have suggested the use of different individual ML classifiers and then apply ensemble methods with some modifications to get accurate results.

III. PROPOSED SYSTEM

Early detection and treatment are the most effective methods for preventing suicidal ideation and attempts. With the increased adoption of mobile, internet, and online social networking technologies, there is a growing trend for people to discuss their suicide plans in online communities. However, this most important social health problem in modern society goes unrecognized, and thousands of individuals commit suicide every year all over the world. As lots of content is being posted on social media by various users it would be better to develop some application to analyze social media posts and to detect the posts with suicidal ideation. The goal of the proposed research is to employ machine learning and deep learning approaches to increase the predictive accuracy of textual post analysis and detect whether a person is having suicidal thoughts via a user interface. It is proposed to develop an API that can be used for detection for of suicidal severity of a social media posts using supervised learning algorithms under machine learning and deep learning domains to detect whether the post contains any suicidal intention.

IV. METHODOLOGY

4.1 Dataset:

The dataset for the implementation contains Twitter posts from users posting suicidal and neutral posts. It consists of total 9119 samples of which 3998 are suicidal posts and 5121 are neutral posts [5].

4.2 Machine Learning approach and implementation:

In the first phase, the proposed system uses natural language processing methods for data preprocessing and feature extraction. For classification, supervised machine learning supervised algorithms like Support Vector Machine (SVM), Naıve Bayes, Stochastic Gradient Descent (SGD), Decision Tree, Random Forest, XG Boost and K Neighbors algorithms. Figure 1 shows the overall methodology using Machine Learning.

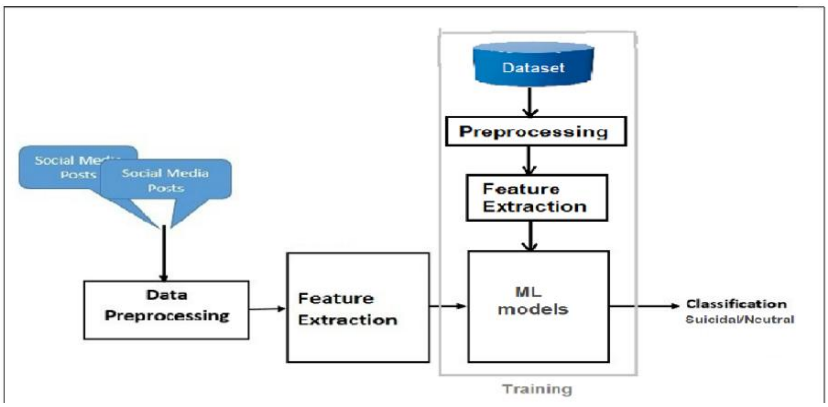

Figure 1: ML approach

4.2.1 Data Preprocessing:

The natural language processing toolkit is the python library which provides various natural language processing modules. Input text data is split into smaller individual components by tokenisation. Stop words are words which do not add significant meaning to the text. Stop words and punctuation are eliminated. Stemming is transforming any word to its most general form or base stem after removing prefixes, suffixes. To reduce vocab without losing valuable information, all words in the tweets are lowercased.

4.2.2 Feature Extraction:

Feature extraction is the process of extracting a list of words from text data and converting it into a set of features. which is to be used for learning. For feature extraction hashing vectoriser is used which converts the input text into a sparse matrix representation having individual token occurence counts. It has a parameter for number of features which is 1048576 by default. The number of features used for experimentation is 2^21.

4.2.3 Supervised machine learning algorithms:

Support Vector Machine (SVM) helps to find a hyperplane (decision boundary) between classification classes by maximizing the margin. The new textual post inputs can be then predicted to belong to a class (suicidal or neutral) based on their respective side of hyperplane. *Naïve Bayes* classifier uses a probabilistic approach based on Bayes theorem. It assumes that each feature is independent of another and hence reduces computational space. *Decision Tree* splits the data into smaller subsets hierarchically based upon the decision evaluation criteria associated with a given variable. It tries to produce the most homogeneous groups at each level. *Random Forest (RF)* uses ensemble approach. It consists of a number of decision trees each of which can be trained upon a random subset of features. It uses majority voting to improve the performance and reduce overfitting. *XG Boost or Extreme Gradient Boosting* uses ensemble technique method that utilizes the results of base learners for making predictions. XG Boost uses gradient boosted decision trees. The parallel tree boosting provided by XG Boost helps to solve text classification tasks in an efficient way. *Stochastic Gradient Descent (SGD)* classifier finds such values of function parameters which minimize the cost function. It calculates the derivative of the loss of a single random data point. Due to efficiency and ease of implementation SGD is used for text classification and natural language processing tasks. *K-nearest neighbor (K-NN)* algorithm classifies by finding the K nearest matches present in the data and then uses the labels of the closest matches to give prediction results. It classifies new data based on similarity. Euclidian distance is prominently used to find the nearest neighbors.

4.3 Deep Learning approach and implementation:

Deep learning finds application in various fields such as computer vision, natural language processing based text categorization, sentiment analysis and medical diagnosis. Word embedding serves as robust feature extraction method to learn complex datasets. In the field of suicide prevention and text analysis of posts deep learning can help identify specific patterns in the data and maintain the context of the words sequentially. Figure 2 shows the overall methodology using Deep Learning.

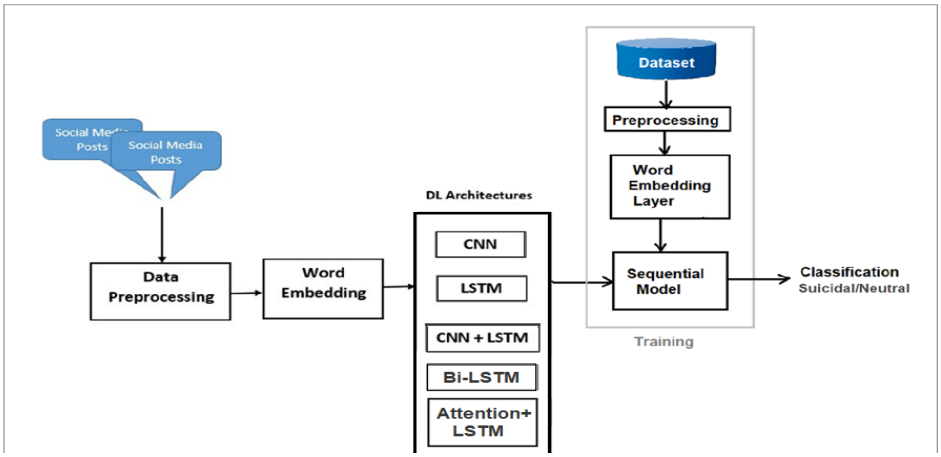

Figure 2: DL approach

4.3.1 Data preprocessing:

For deep learning, cleaning and tokenization of the corpus is done. The corpus is converted to either all uppercase or all lowercase. Interrupting punctuation marks such as full stops, commas and question marks can be represented a s a single special word of the vocabulary. Non-interrupting punctuation marks such as quotation marks can be ignored. Multi-sign marks can be collapsed to a single mark. Numbers can be dropped if they do not carry important meaning in the corpus. Special characters can be dropped. While dealing with tweets special words like emojis and hashtags can be retained with their meanings.

4.3.2 Word Embedding:

Word-embeddings help to use a vector representation in which semantically similar words have a similar encoding. Trainable parameters which are the weights learned by the model during training are used. Keras helps to use Embedding layer. The Embedding layer is much like a lookup table that maps integer indices which represent the words to dense vectors or word-embeddings. The embedding dimension parameter can be set to a value according to the size of the dataset and that which works well with the deep learning architecture After training, the learned embeddings roughly encode the similarities between the words of the textual data.

4.3.3 Convolutional Neural Network (CNN)

CNNs are a type of deep feed forward artificial neural network that uses a multilayer perceptron variation that requires little preprocessing. When we utilise CNN on text data, each convolution's result will fire a trigger when a specific pattern is found. Convolution is the result of applying a filter to an input and getting an activation. When the same filter is applied to an input several times, a feature map is created, displaying the positions and strength of a recognised feature in the input. For textual data, one-dimensional convolutions are utilised. Rectified linear unit (ReLu) activation function is used with the convolution. A pooling layer is another building block of a CNN. Its function is to gradually reduce the spatial size of the representation in order to reduce the number of parameters and computations in the network. Pooling layers work independently on each feature map. As a convolutional network is trained, kernel weights are learned. Each kernel looks at a word and surrounding words in a sequential window and outputs a value. In convolution operation features are recognized as patterns in sequential word groupings and can indicate the sentiment of a text. In one dimensional convolution kernels will slide down a list of word-embeddings in sequence to process an entire sequence of words. Figure 3 shows architecture of CNN model.

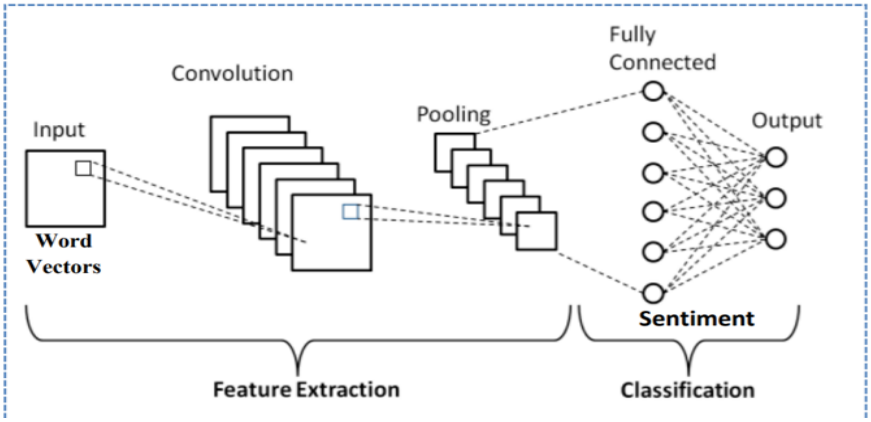

Figure 3: Convolutional Neural Network Architecture

4.3.4 Long Short Term Memory (LSTM)

LSTM is a type of recurrent neural network which maintains the relevant data and discards irrelevant data. LSTM has many hidden layers. While passing through these layers the relevant data is kept and the irrelevant data is forgotten through every single cell. Information gets added and removed through the gates namely input gate, forget gate and output gate. Information from the earlier time steps can be carried to the later time steps thus it remembers patterns and reduces the effect of short memory. In the proposed system LSTM is implemented using the Sequential model. First layer is embedding layer and specification of input shape. Dropout is added to prevent overfitting. Next layer is LSTM layer containing neurons which act as memory unit for the model. Dense layer or fully connected layer is the output layer. Figure 4 shows architecture of LSTM model.

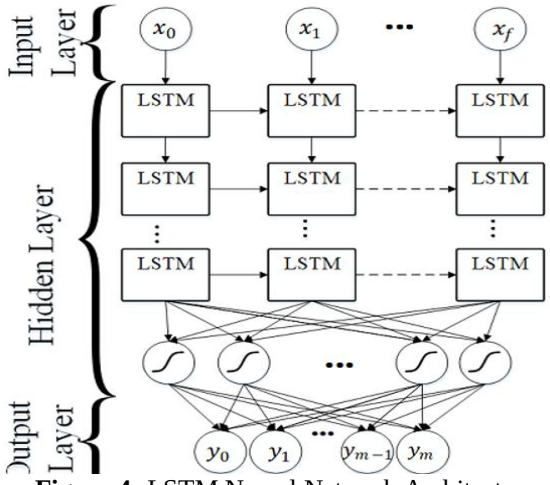

Figure 4: LSTM Neural Network Architecture

4.3.5 Combined LSTM-CNN model

Combined LSTM-CNN architecture consists of the LSTM layer to predict a category for the long term context dependencies and the convolutional layer for extracting important features through pooling which help to identify patterns in the textual data. Figure 4.3.6 shows the LSTM-CNN model for classifying suicidal and neutral textual content. The first layer includes word embedding and specifies the input shape followed by dropout layer, LSTM layer, one dimensional convolutional layer for performing, pooling layer, fully connected layer with an activation function and compilation layer. Figure 5 shows LSTM-CNN architecture.

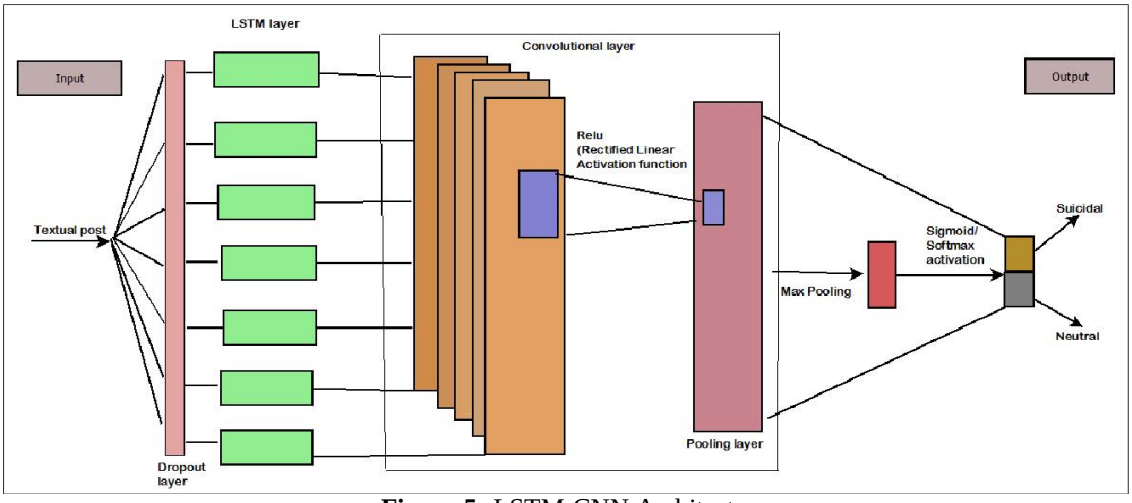

Figure 5: LSTM-CNN Architecture

4.3.6 Bidirectional LSTM (bi-LSTM)

Bi-LSTM neural networks consist of LSTM blocks operating in both directions to combine past and future contextual information. The bi-LSTM network learns long-term dependencies without storing redundant contextual data. bi-LSTM consists of two parallel layers propagating in both the forward and reverse passes to capture the dependencies in the two contexts. Figure 6 shows architecture of bi-LSTM.

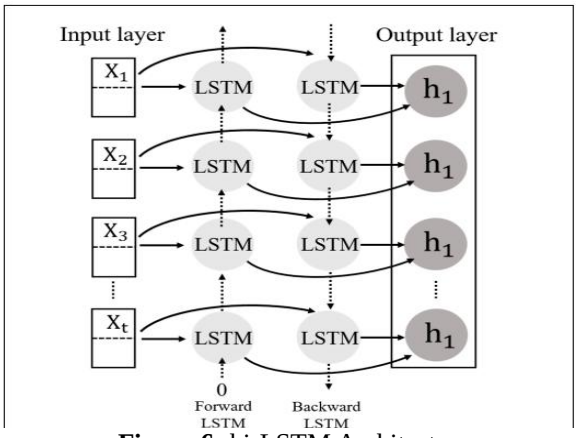

Figure 6: bi-LSTM Architecture

4.3.7 Bidirectional LSTM with Attention mechanism

Attention mechanisms help neural networks focus on specific aspects of complex input data, one at a time, until the entire dataset has been classified. They break the data into smaller areas of attention that can be processed sequentially. Attention mechanism added to bidirectional LSTM improves the classification performance. Bi-LSTM layer processes the sequence of words to obtain information regarding the context between the words. The output coming from each layer is then combined into a self-attention layer. The attention mechanism is used to apply weights to the time steps of the word sequences so that the most useful sub sequences, phrases and words are used to make the prediction. Next is a fully connected dense layer with sigmoid activation that predicts the likelihood of whether the post was written by an author having suicidal intention. Figure 7 shows architecture of Bidirectional LSTM with Attention mechanism.

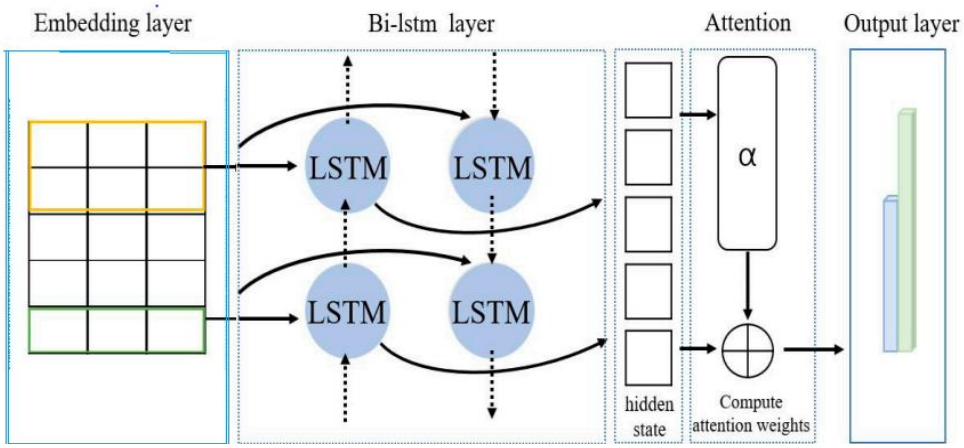

Figure 7: bi-LSTM with Attention mechanism

4.3.8 Parameter Settings

Some of the comparable parameters of the deep learning architectures as used for the deep learning architectures in the text classification task are given in Table 1. Apart from the parameters mentioned in the table some of the settings were such: for CNN model the max pooling size was 2 and it was trained for 15 epochs, single layer LSTM and 2 layer LSTM had a spatial dropout 0.7 and were trained for 30 epochs, LSTM-CNN model had pooling size as 5 and was trained for 10 epochs, bi-LSTM and Attention-bi-LSTM had spatial dropout of 0.7 and were trained for 10 epochs. All the models used adam as optimizer and binary crossentropy as loss function.

Table 1: Deep learning models parameter settings

Models	Batch size	LSTM cells	No. of filters	Windows	Conv1d activation	Fully connected layer activation	drop-out
CNN	64	-	128	3	ReLu	sigmoid	-
LSTM	500	200	-	-	-	sigmoid	0.2
2 layer LSTM	500	80, 40	-	-	-	sigmoid	0.2
LSTM-CNN	504	100	128	3	ReLu	sigmoid	0.5
Bi-LSTM	128	64	-	-	-	sigmoid	0.5
Attention Bi-LSTM	32	64	-	-	-	sigmoid	-

V. RESULTS

5.1 Machine learning performance and results

The various performance metrics for Machine Learning models are as shown in the Table 2.

Table 2: ML models performance

Models	Precision	Recall	F1-score	Accuracy %
SVM	0.94	0.87	0.90	92.0
Bernoulli NB	1.00	0.10	0.18	61.1
Decision Tree	0.87	0.86	0.87	88.5
Random Forest	0.94	0.84	0.89	90.7
XG Boost	0.92	0.90	0.91	92.4
SGD	0.93	0.87	0.90	91.3
K-NN	0.77	0.20	0.32	62.9

Figure 8 shows comparison of accuracies of different Machine Learning algorithms.

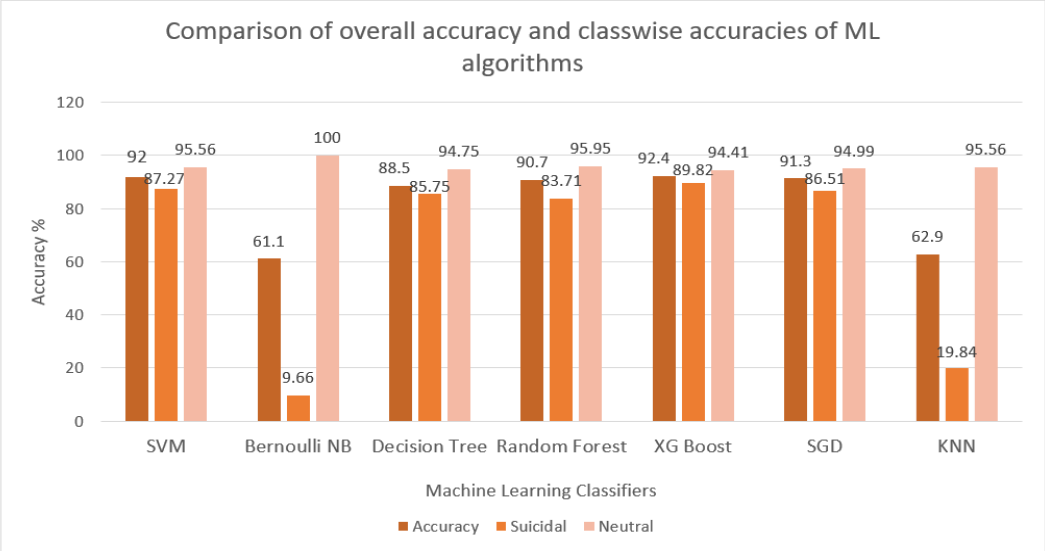

Figure 8: Overall and class-wise accuracies of ML models

Upon comparing different ML models, it is observed that the best overall classification is achieved with XG Boost classifier. It shows 89.82 % accuracy for suicidal posts which is the highest for suicidal class as compared to other models and 94.41% accuracy for neutral posts. XG Boost classifier significantly outperformed the other models and demonstrated highest overall predictive accuracy of 92.4%.

5.2 Deep learning performance and results

The various performance metrics for the different deep learning models are shown in the Table 3.

Table 3: DL models performance

Models	Precision	Recall	F1-score	Accuracy %
LSTM(single layer)	0.89	0.93	0.91	91.88
LSTM(2 layers)	0.91	0.92	0.91	92.43
CNN	0.87	0.91	0.89	90.57
LSTM-CNN	0.965	0.86	0.909	92.65
Bi-LSTM	0.946	0.895	0.92	93.31
Attention-bi-LSTM	0.942	0.91	0.926	93.75

Figure 9 shows comparison of accuracies of different deep Learning algorithms.

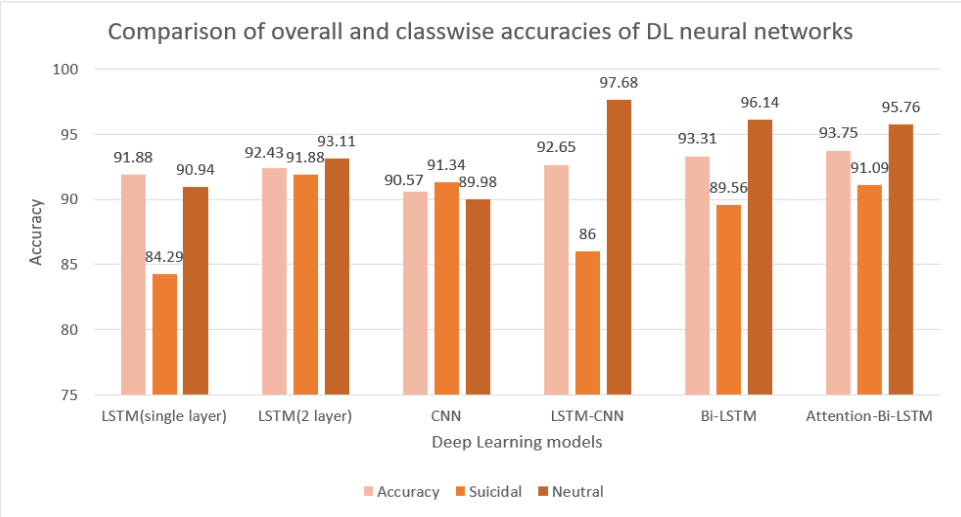

Figure 9: Overall and class-wise accuracies of DL models

Upon comparing different DL models, it is observed that the best overall classification is achieved with

Bi-LSTM with attention mechanism. It shows 91.09 % accuracy for suicidal posts and 95.76 % accuracy for neutral posts. With the optimized parameters, Attention Bi-LSTM model significantly outperformed the other models and demonstrated highest overall predictive accuracy of 93.75 %. Considering the suicidal class accuracies of various DL models it is found that LSTM (2 layers) is having maximum suicidal class accuracy of 91.88%.

VI. CONCLUSION

Early detection and treatment is one effective way to prevent depressed individuals from falling prey to suicide attempts. Social media platforms and online communities act as powerful tools to identify at risk individuals and extend help to these individuals. It serves as a medium for socially active groups and mental health caretakers to identify and help users exhibiting suicidal ideation over such networks. Machine learning and Deep learning classifiers are trained over Twitter dataset and used for the text classification task to categorise social media posts into suicidal or neutral category. Attention-bi-LSTM exhibits highest predictive accuracy 93.75 % hence it is the most suitable deep learning neural network architecture for prediction. In this project, an API is proposed that can be used to tag any textual data with a potential suicidal thought tag or neutral tag. The API is built using flask server connected to the pre-trained models. This was an attempt to create a classification system that can be utilised with the help of suitable modifications so that the generalised social media user content can be categorised and examined for the presence of any kind of suicidal tendencies. Addressing the important social issue of suicide attempts among individuals beforehand by leveraging the context of the posts made by people is the main inspiration of this project. The work can be extended to find the varying levels of mental health concerns exhibited by social media content creators by introducing few more classification classes over the data. In this new age of connectivity where online textual content is generated every minute, the scope of this project is not only limited to suicidal thoughts prediction but also this data can be leveraged in many ways like deep sentiment analysis, healthcare-related problems can be solved, anxiety and stress levels among individuals can be monitored, hate and toxicity of the social media posts can be detected to stop bullying, etc.

REFERENCES

[1]. M.Mesfin, H.Lin, and B.Xu ,"Detection of Suicide Ideation in Social Media Forums Using Deep Learning," MDPI Algorithms, Vol.13, December 2019. Available: https://www.mdpi.com/1999-4893/13/1/7.

[2]. I. Aminu, "Building a suicidal tweet classifier using NLP," Medium, 18-Aug-2020. [Online]. Available: https://towardsdatascience.com/building-a-suicidal-tweetclassifier-using-nlp-ff6ccd77e971.

[3]. S. Ji, S. Pan, X.Li, E.Cambria, G.Long, and Z. Huan," Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications," IEEE Transaction on Computational Social Systems,Vol.8, September 2020.

[4]. S.Dabhane, and M.Chawan, "Depression Detection on Social Media using Machine Learning Techniques: A Surve," International Research Journal of Engineering and Technology (IRJET), Vol.7, November 2020. Available: https://www.irjet.net/archives/V7/i11/IRJET-V7I1116.pdf.

[5]. AminuIsrael, "AminuIsrael/predicting-suicide-ideation: A notebook on building a suicide ideation classifier using natural language processing(nlp)," GitHub,27-Apr2022 [Online]. Available: https://github.com/AminuIsrael/Predicting-Suicide-Ideation.