Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14431-z

Stacked CNN - LSTM approach for prediction of suicidal ideation on social media

Bhavini Priyamvada[1] · Shruti Singhal[1] · Anand Nayyar[2,3] · Rachna Jain[4] · Priya Goel[1] · Mehar Rani[1] · Muskan Srivastava[1]

Received: 15 July 2021 / Revised: 15 May 2022 / Accepted: 22 January 2023

Abstract
The growing use of social media forums to express suicidal ideation creates an immense requirement for automatic recognition of suicidal posts. Individuals use social forums to discuss their problems or access information on similar topics. This study aims to work on the automatic recognition and flagging of suicidal posts. It presents an approach that analyses social media platform Twitter to identify suicide warning signs for individuals. The primary purpose of the mentioned approach is the automatic identification of abnormal changes in online behaviour of the user. The challenges faced in suicide prevention is the understanding and detection of complex risk factors or warning signs that may lead to the event. To achieve this task, numerous natural language processing (NLP) techniques are employed to quantify linguistic and textual changes and pass through a novel framework which can be applied at large. The preliminary detection of suicidal ideation is achieved through deep learning and machine learning-based classification models applied to tweets on Twitter social media. For both classifiers initially we executed data pre-processing, feature extraction subsequently machine learning and deep learning classifiers respectively. For this purpose, we employ a Stacked CNN - 2 Layer LSTM model to evaluate and compare with other classification models. The study shows that the Stacked CNN - 2 Layer LSTM architecture with word embedding techniques achieves 93.92% classification accuracy as compared to previous previous CNN - LSTM approaches.

Keywords Early suicide detection · Suicide ideation · Word embeddings · Machine learning · Deep learning · Twitter

✉ Anand Nayyar
anandnayyar@duytan.edu.vn

Extended author information available on the last page of the article

Published online: 24 February 2023

1 Introduction

Suicide attempts and suicide are significant individual's health issues. There could be many aspects which can cause suicidal ideation. Several individuals fall victim for showing suicidal tendencies each and every year. Consistently, just about 800,000 individuals commit suicide. Suicide stays the subsequent leading reason for death among a young generation with an overall suicidal rate of 10.5 per 100,000 individuals. It is anticipated that by 2020, the death rate will increase to one each 20 s. Almost 79% of the suicides happened in low and middle income nations where the resources for the identification and management are often scant and insufficient [47]. There is a need in progression of suicidal detection technologies by examining social media content. Preceding studies and research have demonstrated that adolescents are more probable to communicate suicidal feelings on social media platforms. Also, suicidal thoughts may not be shared with physicians. It is indeterminate to what scope the online thoughts are corresponding to suicidal behaviour as induced by the physicians. Most of the suicidal attempts can be intercepted and it's vital to know and understand the process through which people convey their thoughts and feelings.

Suicidal ideation is seen as an inclination to take ones' life ranging from depression, through a plan for a suicide attempt, to an intense preoccupation with self-destruction [3]. At risk, individuals can be perceived as suicide ideators (or planners) and suicide attempters (or completers) [40]. As per a few studies, the majority of suicide ideators do not make attempts to commit suicide. Klonsky and May [23] believed that most cited risk indicators or factors (depression, hopelessness, frustration) associated with suicide are the indicators of ideation and do not lead to an attempt. On the Contrary, Pompili et al. [35] uncovered that both the ideator and the attempter might be similar to many variables categorised as risk factors for suicidal behaviour. In WHO nations, early identification of suicidal ideation has been developed and carried out as a national suicide prevention strategy to pursue towards the worldwide market with the common intent to reduce the suicide rates by 10% [48].

In recent years, online platforms have become a window into their users' psychological wellness and prosperity, mainly the youth generation. It offers anonymous participation in various cyber communities to give a space for a public conversation about socially stigmatised subjects. The content generated and published on these platforms often indicates the well-being of individuals. Any written sign about suicide can be viewed as a stressful indicator; therefore, addressing these claims and flagging content when published is an additional safeguard. It also protects the audience who doesn't have any suicidal behaviour from any triggering information which may negatively impact individuals. Social media texts, such as blog posts, forum messages, tweets, and other online notes, are recorded in the present and are well preserved, limiting any misleading interpretations produced by retrospective analysis [11].

Mental well-being has become a forefront of various forums online, becoming an arising zone to analyse in computational linguistics. It provides a powerful research platform for advancing new innovative technological methodologies and enhancements, bringing a novelty in suicidal detection and further suicide risk prevention [27]. Ji et al. [21] have proposed a novel content protecting solution and advanced optimisation strategy (AvgDiffLDP) for early identification of suicidal ideation. Their solution was to train a local data-preserving model for every local user that shares their own model parameters with the server in spite of an individual's personal information. They have developed an average difference descent algorithm, which is a novel updating algorithm which combines parameters from various client models to optimize the learning capability of the model. The techniques used earlier were just

based on Machine Learning which has their limitations such as negative transfer learning and overfitting of the dataset.

We observed that apart from conventional text classification approaches, deep learning techniques have already made a noteworthy development for pattern recognition and computer vision. Neural networks based on vector representations can produce superior outcomes on different Natural language processing (NLP) tasks contrary to the traditional machine learning approaches [50]. In combination with word embeddings [28, 29], deep neural networks outperform conventional machine learning algorithms for suicidal risk assessments. This research aims to differentiate the various machine learning and deep learning classifiers, thereby developing a feasible solution. Additionally, it expects to expand the present flagging technology by adding and highlighting tokens that associate more with suicidal and non-suicidal labels. By exploring the potential of Convolutional Neural Network (CNN) [24], Long Short-Term Memory (LSTM) [19] and their combined model, the study demonstrates that Stacked CNN - 2 Layer LSTM combined model can outperform the performance of individual CNN or LSTM model, and traditional machine learning techniques for suicide-related topics.

1.1 Main contributions of the research

- To utilize the n-gram analysis on certain suicide-related forums that express suicidal tendencies along with reduced social engagements mentioned in these forums to identify the transition towards suicidal ideation often associated with different stages such as self-focused attention, hopelessness, anxiety, frustration, or loneliness.
- To assess the TF-IDF and performance of statistical features over word embeddings using CNN, LSTM and Stacked CNN - 2 Layer LSTM model analysis.
- To explore the performance of Stacked CNN - 2 Layer LSTM combined class of deep neural networks as the proposed model for identification of suicide ideation to improve the existing method and to evaluate its potential with CNN-LSTM, CNN, LSTM deep learning techniques and traditional machine learning classifiers on the real-time dataset.
- To compare Stacked CNN - 2 Layer LSTM model with previous CNN-LSTM models to validate the model accuracy.

1.2 Organization of paper

The paper is organized as: Section 2 describes background and related work on suicide and suicide ideation detected in social media. Section 3 highlights materials. Section 4 highlights proposed methodology. Section 5 focusses on baseline, model architectures and its parameters and evaluation metrics. In section 6, enlightens the results and analysis for detection of suicide ideation, and section 7 concludes the paper with future scope.

2 Background and related work

In recent times, various researchers have shown interest in understanding patterns on social media, especially those demonstrating suicidal behaviour. Tadesse et al. (2020) [44] developed an incorporated LSTM-CNN model in order to assess and contrast with more classification models. Their research showed that the integrated neural network model along with word

embedding techniques could accomplish the best appropriate classification outcomes. Moreover, their outcomes assist the strength and capacity of deep learning models to develop a viable model for suicide ideation assessment in several text classification tasks. Sun et al. (2019) [43] developed an additional text approach from this feature and changed ABSA into a sentence-pair classification task. They attained new state-of-the-art outcomes on SemEval-2014 Task 4 and SentiHood datasets by modifying the pre-trained model from BERT. Munikar et al. (2019) [31] utilised a promising machine learning model called BERT for addressing the fine-grained sentiment classification task. They pre-trained BERT model and utilised it for the fine-grained sentiment classification on the Stanford Sentiment Treebank (SST) dataset. Experiments demonstrated that their model beat various other models without complex architecture. Likewise, they demonstrated the viability of transfer learning in Natural Language Processing. Nordin et al. (2021) [32] used feature selection algorithms to provide a comparative study on single predictive models along with ensemble predictive models for comparing between individuals who attempt suicide from those who do not. They analysed a dataset comprising of 75 patients having depression and implemented various machine learning algorithms. The result showed that ensemble predictive models surpassed the single predictive models. The highest accuracy was achieved by voting and bagging in contrast with various machine learning algorithms.

Sawhney et al. (2018) [37] contrasted the accomplishment of three deep learning models, specifically CNN, RNN, LSTM with standard baselines in text classification problems. In this, a lexicon of words and phrases having a link with suicidal text was generated by getting known suicide web forums for dataset formation. The dataset comprised of 300 posts from online suicide forums and 2000 arbitrarily collected posts from Reddit and Tumblr. The efficacy of the C-LSTM model for suicidal ideation assessment in tweets was obtained through quantitative comparison among the models. The ability of CNNs to spatially encrypt tweets into a 1-D structure to be fed into LSTMs, and the capability of LSTMs to catch long-term dependencies were the two factors that helped the model succeed. Ji et al. (2018) [20] analysed language preference of users and topic descriptions as a warning system to detect suicidal ideation. Their study contrasted six classifiers, which included four traditional supervised classifiers and two neural network models. For evaluation of suicidal ideation, they extracted various descriptive sets of syntactic, statistical, linguistic, topic features and word embedding. The dataset extracted from Reddit and Twitter comprised of suicide ideation texts. The study provided substantial insight which can complement the comprehension of suicidal ideation.

Chadha and Kaushik (2019) [6] pre-processed tweets in accordance with the semantics of recognized features and then transformed them to probabilistic values. This facilitated the implementation in machine learning and ensemble learning techniques. Various Machine Learning algorithms such as Multinomial Naive Bayes, Bernoulli Naive Bayes, Decision Trees were also tested on the data to analyse suicidal ideation. Kumar et al. (2020) [36] studied productive methods to deter individuals from suicidal ideation and suicidal attempts. Suicidal ideation via sentiment analysis and supervised learning methods were implemented on online user content like Twitter data, for early detection in their research. Lin et al. (2020) [26] utilised machine learning techniques like Support Vector Machine, Decision Tree, Logistic Regression Tree and Multilayer Perceptron to prognosticate suicide ideation with the help of some necessary psychological stress zones of the military males and females. The result obtained showed that the accuracies of the implemented machine learning algorithms were above 98%. Multilayer Perceptron and Support Vector Machine obtained the best predictors of suicide ideation reached up to 100% accuracy.

Sawhney et al. (2018) [38] proposed a set of features in order to train ensemble and linear classifiers for a dataset consisting of physically annotated tweets. Baseline models using several strategies like Negation Resolution, LSTMs, Rule-based methods were compared in this research.

Experimental results showed the enhanced performance of the Random Forest classifier when contrasted to other classifiers as well as the baselines. Evidence suggested that the feature extraction accompanied by the linear and ensemble classifiers succeeds the baselines presented in terms of performance. Kalchbrenner et al. (2014) [22] supported the strength of CNN on n-gram features from numerous text positions. Yin and Schütze (2016) [51] proposed an unsupervised pre-training and multichannel word embedding model, enhancing classification accuracy. Morales et al. (2019) [30] presented the CNN and LSTM models for suicidal ideation depicting the results for a tested personality and tone features. Gehrmann et al. (2018) [14] employed LR and cTAKES techniques along with n-gram features for comparison of CNN model with traditional rule-based entity extraction systems, which showed CNN performed better than other classifying approaches.

Bhat and Goldman-Mellor (2017) [4] foregrounded results and performance of CNN over various other techniques to detect suicidal behaviour. Du et al. (2018) [12] utilized deep learning techniques for the prediction of suicidal tendencies and formed a binary classifier to differentiate between suicidal and non-suicidal content on social media. He and Lin (2016) [18] presented a neural network model formed on a hybrid of BI-LSTMs and ConvNet, which contributed to resolving the measurement problem of semantic textual similarity. Sinha et al. (2019) [41] employed a multipronged technique and used various neural network models trained on content posted on Twitter. They trained a stacked ensemble of classifiers constituting contrasting forms of suicidal behaviour. They examined numerous trained models and provided an assessment depicting how historical tweeting and information stored in the homophile networks amidst users in Twitter contributes in precisely distinguishing tweets expressing suicidal purpose.

Zhu et al. (2020) [56] used text mining to screen suicidal ideation. Several combinations of two-term weighting factors and six algorithms were used and contrasted for their performance under different training set sizes. The results indicated that AdaBoost, Random Forest weighted by TF and SVM showed better generalisation ability and were able to filter out suicidal patients with a slight quantity of representative terms. Choi et al. (2018) [8] explored the possibility of suicide-death with the help of baseline characteristics and medical visit history data and used Support Vector Machines (SVMs), cox regression and Deep Neural Networks (DNNs). Weng et al. (2020) [46] presented a model formed on machine learning and 3D autoencoder, which predicted individuals demonstrating suicidal ideation by referring to their structural brain imaging. They performed their research on the GQI dataset where they trained indices of isotropic values of the orientation distribution function (ISO), generalised fractional anisotropy (GFA) and normalised quantitative anisotropy (NQA) on different machine learning models. The GQI dataset incorporated various groups, namely 58 healthy controls (HC), 54 depressive patients without any suicidal ideation (NS) and 41 depressive patients showing suicidal ideation (SI). The results obtained from their research indicated that the best pattern of structure across multiple brain locations could distinguish SI from HC and NS, achieving good results. Li et al. (2021) [25], developed a multifeatured fusion recurrent attention model for suicide risk assessment and used bidirectional long short-term memory network to create the text representation with context information from social media posts. They further introduced a self-attention mechanism to extract the core information. They have also used fused linguistic features to improve their model.

The previous research approaches discussed here all made use of either machine learning or deep learning models along with pre-trained word embeddings to improve the precision of their models. The dataset's size previously used was not enough to provide conclusive suicidal ideation detection methodology. All of these models have their separate set of limitations like negative transfer learning and overfitting on the dataset. Hence, instead of using a single one of the deep learning architectures, several of these neural network models can be combined to cover the problems that they may face

Table 1 Comparing previous works on Prediction of Suicidal Ideation

Author	Model	Features	Dataset	Accuracy (in %)
Tadesse et al. [44]	LSTM-CNN	Word2vec	Reddit	93.8
Sawhney [37]	C-LSTM	Word2vec	Twitter	81.2
Ji [20]	LSTM	Word2vec	Reddit	92.66
Sinha [41]	Bi-LSTM	Attention Layer	Twitter	92.20
Proposed Model	Stacked CNN - 2 Layer LSTM	Word2vec	Twitter	93.92

when other researchers use them separately. By observing the past work and in order to enhance our model, we have used the dataset from another popular social media platform namely Twitter dataset. As shown in Table 1 above, we can observe that in the past, there were not any studies using the Stacked CNN - 2 Layer LSTM models for suicidal ideation detection, by using this combined neural network model we were able to overcome the limitations such as overfitting and underfitting of the individually trained models. The Stacked CNN - 2 Layer LSTM model used in our study not only improved stability but also accuracy and predictive power of the proposed model.

3 Materials and methods

3.1 Dataset description

We have collected our dataset from GitHub website [15] which consists of various tweets from the Twitter website in order to detect suicidal behaviour on social media. We took Twitter's tweet data which consists of suicidal intention and non - suicidal intention data. The dataset consists of Tweets representing the text data and intention representing the suicide ideation label. The value 1 indicates that tweet has suicidal ideation and the value 0 indicates that tweet does not have suicidal ideation. The dataset used consists of approximately 10,000 tweets to further identify the potential suicidal tweets. Among 10,000 tweets 5126 are classified as suicidal ideation, whereas 4833 were classified as non-suicidal. Our dataset consists of approximately equal distribution of suicidal and non-suicidal tweets. Here, Figs. 1 and 2 represents the word cloud for the most frequently used words in non - suicidal and suicidal tweets.

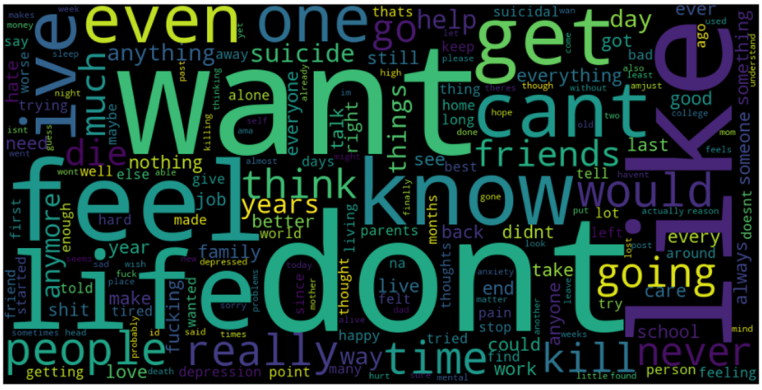


Fig. 1 Word cloud for Non-Suicidal intention

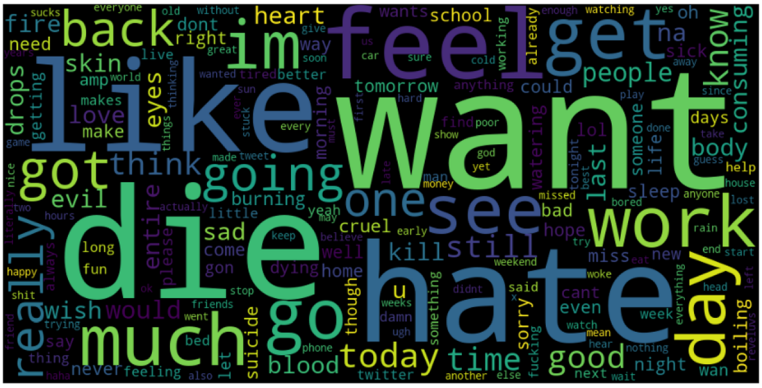


Fig. 2 Word cloud for Suicidal intention

4 Methodology

The motive of our study is to use and execute various machine learning and deep learning classifiers to predict suicidal ideation in social media and compare their outcomes so as to devise the best model for suicidal intention detection. We implemented a combination of deep learning techniques to enhance the performance of language modelling and text classification to identify suicide ideation on twitter website. Figure 3 shows an outline of our proposed framework. It comprises two directions for text data mining methods. The first framework comprises data pre-processing, features extraction with NLP techniques (TF-IDF) utilized to encode the text to additionally continue by traditional machine learning systems for the baseline methods. The second one is produced by data pre-processing, features extraction using word embedding (Word2Vec), subsequently deep learning classifiers for the proposed model. We will provide a comparative analysis using various evaluation metrics such as precision, accuracy, recall and F1-score for the baseline methods and proposed model, hence obtaining the best model for the task of suicide ideation detection.

4.1 Data pre-processing

Data pre-processing incorporates filtering of content in order to get an improved version of content (e.g., Suicidal text) which helps in enhancing the accuracy of the model. A series of filters are applied on the text data. We made use of the Natural Language Toolkit (NLTK) [5] to filter content in advance to the training stage. Several steps have been employed to pre-process the text involving concatenation of text, removal of duplicate words, implementation of tokenization for individual tokens, replacing unnecessary text with a single whitespace, removal of content (e.g., stopwords, colons) and lemmatization to get lemma of text after performing its morphological analysis.

4.2 Stacked CNN - 2 layer LSTM model

a) System Model

For suicide ideation detection, we incorporated the power of both CNN and LSTM neural network and employed an integrated Stacked CNN - 2 Layer LSTM model text classification. The output vector of CNN is passed as an input vector of LSTM to the

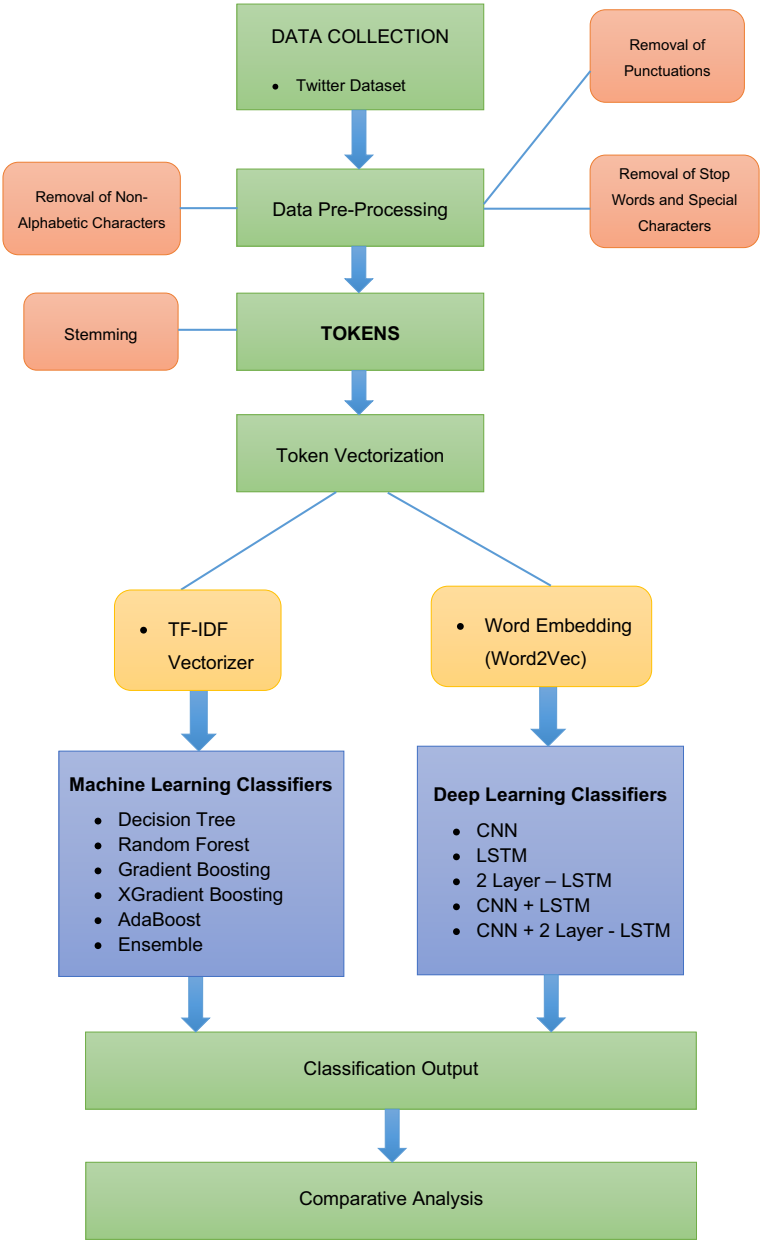


Fig. 3 Suicide Ideation Detection Framework

model and again the output of CNN - LSTM model combined is passed onto the input of another LSTM layer, which helps our model to form a new LSTM model on the CNN to get the features of the content and also to achieve better results.

We utilized the Hybrid framework for Text modelling utilizing the CNN and LSTM combined method implemented in preceding studies [1, 42, 55]. Figure 4 depicts the

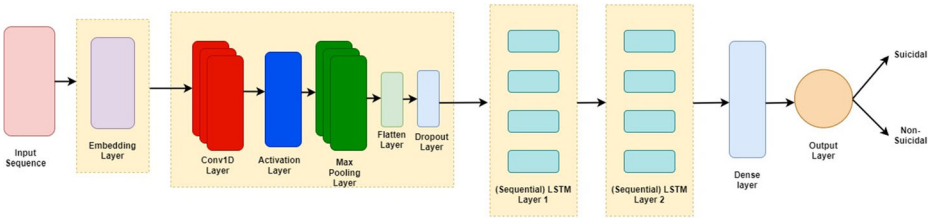


Fig. 4 Suicide Ideation Detection Architecture

proposed Stacked CNN - 2 Layer LSTM combined model architecture for classifying the tweets with suicidal and non-suicidal ideation.

First layer of our model is a word embedding layer, in this a fixed length vector is formed by assigning a unique index to every word in a given sentence. Word embedding layer is accompanied by convolution layer which implements feature extraction followed by Activation, Max pooling, Flatten, and dropout layer which helps in avoiding overfitting. Following these layers, a layer of LSTM is applied to extract long distance reliance across the content complemented by the next LSTM layer, the pooling layer combines the information extracted in order to pool a feature dimension and then it is transformed to a column vector with the help of a flatten layer. Softmax performs a classification which completes the entire neural network process.

b) Architecture

Word embedding layer Word embedding [33] is the representation of text into numerical vectors. Being an input layer of Stacked CNN - 2 Layer LSTM, it maps the textual words into a real number vector space. We utilized word2vec which categorizes the vector of related words together within a vector space on the basis of mathematical similarities by training two neural layers [29]. Text comprising of series of words x_1; x_2; x_3; ...; x_T is transformed into word vectors by pretrained word2vec [29] which helps in characterizing by index number of embedding layers by converting this sort of indices to the d-dimensional of embedding vectors $X_t \in R^d$. In the mentioned expression, d represents the dimension of word vector, accompanied by an input text given in equation (1)

$$X = [x_1; x_2; x_3; \ldots; x_T]^{T \times d} \tag{1}$$

The *t*th word is represented as $X_t \in R^d$, t represents text length and d represents dimension of the word vector. We utilized the openly accessible Word2Vec vectors, which are pre-trained on 100 million words available on Google News having 300 dimensional vectors,

4.2.1 Convolutional layer

Convolutional layer is the central component of the Convolution Neural Network [24]. It contributes to multiple text classification tasks [22, 53, 54]. The model will find and

learn patterns by applying CNN on highly organized content. Convolutional layer comprises various filters of which parameters are to be learned. Every filter is convolved with input volume for calculating a feature map consisting of neurons. The feature maps of each and every filter are stacked along the depth dimension to get the resultant volume. The connections in CNN do not form a cycle just like a feed-forward network. An individual neuron in CNN indicates a region in the input sample.

Since, the feature map is acquired by executing convolution among input and filter, the filter parameters are shared for each and every confined position. Following the extraction of each feature from the embedding layer. We utilized numerous convolutional filters accompanied by various parameter initializations in order to get various maps from the content.

4.2.2 Long short-term memory

LSTM [19] is a sequential network which lets information to persevere. It is an advanced version of Recurrent Neural Network (RNN). LSTM networks are a kind of RNN network which makes use of special units as well as standard units. It comprises a memory cell which is responsible for the course of information from each and every gate which helps to find suicidal content. LSTM helps in prevention of explosion or vanishing gradient frequently perceived in RNN models [37]. We implemented and applied a double layer along with 100 LSTM and 50 LSTM units. For each and every cell, four independent computations were carried out with the help of four gates. LSTM layer formation with input sentences $X = (x_t)$ accompanied by a d-dimensional word embedding vector, H represents the LSTM hidden layer nodes [55].

4.2.3 Pooling layer

The pooling layer is employed to decrease the spatial dimensions excluding depth on a convolutional neural network which assists in preserving essential information. It makes the input depiction smaller and feasible collecting useful data. It helps to decrease the number of calculations and parameters managing over-fitting [49]. Max pooling operation is enforced by us to get essential information in every feature map.

4.2.4 Flatten layer

Flatten Layer modifies a pooled feature map to a column vector which is passed as an input to the neural network classification task [1]. The pooled feature maps are flattened by a reshape function.

4.2.5 Output layer

Output Layer computes the probability of suicidal ideation text with the help of text feature vector from convolution layer, output of pooling layer and activation functions. On our output layer, we used the SoftMax [16] function for classification of input text to binary classification.

5 Baseline, Model Architecture and Parameters and Metrics

5.1 Baseline

In our baseline model, features such as TF-IDF, Bag of Words and Statistical Features were extracted from the text and later passed to machine learning approaches, namely Decision Trees, Random Forest, Gradient Boosting, XGradient Boosting, Adaptive Boosting, Ensemble techniques.

TF-IDF is a method that is used to calculate the weight for every word. This expresses the value of the word in the document and corpus. It ensures text analysis only includes important words by assigning a higher value to more relevant words and does not select the words with lower importance [45].

Decision Trees are a supervised learning method that learns easy decision rules from the data provided and predicts the value of the target variable. Decision Trees were the framework used to support our model using statistical tests. In combination with AdaBoost, they were able to detect dependencies and patterns among classes.

Random Forest (RF) algorithm is an ensemble method employed to combine weak classifiers to make a robust classifier implemented and utilised to solve binary class classification problems [13]. We trained a random forest model using neural network outputs to predict binary suicide ideation status [39].

An Ensemble of the above machine learning classifiers was used to generate a single optimal predictive model. This helped lower the spread of predictions in our study.

Adaptive Boosting (AdaBoost), Gradient Boosting and XGradient Boosting (XGBoost) algorithms were used to produce a combined prediction model by forming an ensemble of weak prediction models (decision trees). This helped recognise patterns in residuals which contributed to making our model strong and better. XGBoost [7] algorithm provided controlled model formalisation in controlling over-fitting.

The objective of using Machine Learning classifiers was that these algorithms can distinguish better between text written by individuals who have died by suicide in contrast with simulated suicide notes when compared by professionals in mental health by a score of 71% vs 70% [34]. Moving forward, we have contrasted the above ML models with pre-defined Deep Learning models like CNN and LSTM to compare the accuracy and improve the existing flagging system, thereby helping automate the process of detection in suicidal ideation.

5.2 Model architecture and its parameters

In this study, we trained our Stacked CNN - 2 Layer LSTM combined model on its earlier execution for classification. Fine-tuning with 10-fold cross-validation was performed. We applied a pre-trained word2vec model which was trained on 100 billion words from Google News for features classification. The neural network models are initialized with a 300-dimensional pre-trained word2vec [9, 29].

Table 2 presents the parameter setting for the proposed model (Stacked CNN - 2 Layer LSTM). The experiment is conducted using different parameters listed as follows: the parameters, namely number of filters, kernel size, padding, pooling size, embedding dimensions, activation function, optimizer, dropout, batch size, epochs and units. We used Python with the NLTK natural language toolkit.

Table 2 Hyperparameters used in the Proposed Model

PARAMETERS	VALUES
Filters	64
Kernel	3
Padding	Same
Embedding Dimension	300
Epochs	20
Activation Function	ReLU
Loss	Binary Cross entropy
Batch Size	64
Word Embedding	Word2Vec
Max Pooling	2
Dropout	0.5
Optimizer	Adam
Fully connected Layer	SoftMax

5.3 Evaluation metrics

To appraise our proposed model with the baseline, we made use of evaluation metrics consisting of computing Accuracy, Precision (P), Recall (R) and F-score (F1) given by the following equations. It depends upon confusion matrix which includes details regarding every test sample prediction result. Accuracy is the ratio of correct number of predictions to the total number of predictions to know how much precise our proposed model is; Precision (P) estimates the number of positively identified samples; Recall (R) approximates the proportion of correctly identified positive samples; F-score (F1) is the harmonic mean of recall and precision. is. In the evaluation metrics, we calculate the number of true positive predictions (TP) in which the model accurately detects the positive class, outcome in which the model accurately detects the negative class is true negative predictions (TN), outcome in which model inaccurately detects the positive class is false-positive predictions (FP) and the result in which the model inaccurately detects the negative class is false-negative predictions (FN) [2]. The classifying evaluation score is an accuracy given as following:

Here, Accuracy is the ratio of the individuals who are truly either attempting suicide or not, to the total number of individuals considered.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision or positive predictive value is the probability that following positively classified cases, that individual will truly attempt suicide.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall or true positive rate is the proportion of individuals who tests positive among all those who are actually attempting suicide.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F1- Score is the harmonic mean of individuals who will truly attempt suicide and the individuals who tests positive among those who will actually attempt suicide.

$$F_1 = 2\frac{Precision\ .\ recall}{Precision + recall} \tag{5}$$

Some errors that we have faced during our study:

- False negatives: suicidal tweets not being detected by the model, most probably because they are very specific to a certain situation or culture and they require a high level of world knowledge that Machine or deep Learning models don't have. The most effective sarcasm is the one tailored specifically to the person, situation and human variability.
- Suicidal ideation tweets written in a very polite way are undetected. Sometimes people use politeness as a way of hiding their suicidal ideation, highly formal words that don't match the casual conversation.

6 Results analysis

In this section, we will talk about the different approaches that we have applied to the dataset and the result acquired by it. We arrived at the results in two main steps. In the beginning, data analysis results in the whole collection of tweets were examined. Firstly, the most recurrent n-grams in suicide-indicative posts were analysed and compared with the n-grams in non-suicidal posts. After that, the proposed set of features were used in order to compute the sign of suicidal thoughts and the performance of our suggested deep learning model along with the baseline model was compared with the help of evaluation metrics mentioned above.

6.1 Data analysis results

We examined the entire dataset in order to explore the existence of suicidal thoughts to differentiate the dissimilarities in the lexicon. We took tweet data of suicidal intention and no intention data in order to detect suicidal behaviour in social media content. Tweet represents the text data and intention represents the target level. A visual support of the word cloud was used. The word clouds for non-suicidal intention and suicidal intention are shown in Figs. 1 and 2 respectively. The value 1 indicates that tweet indicates suicidal behaviour and the value 0 indicates that tweet does not indicate suicidal behaviour. A series of filters were applied on the text data. We used the Natural Language Toolkit (NLTK) to filter content in advance to the training stage. Several steps were employed to pre-process the text involving concatenation of text, removal of duplicate words, implementation of tokenization for individual tokens, replacing unnecessary text with a single whitespace, removal of content (e.g., stopwords, colons) and lemmatization to get lemma of text after performing its morphological analysis.

A very high number of question marks ("Why is mankind afraid of death?", "Concerned but don't know what to say?") were observed in the suicide intention tweets. Additionally, a sense of hopelessness can also be seen in the data. Contradicting the suicide intention tweets, the data studied in the non-suicidal tweets accommodated primarily the words expressing positive attitude, joyous moments, and feelings ("joke", "laugh", "want fun", "want happy").

6.2 Classification analysis results

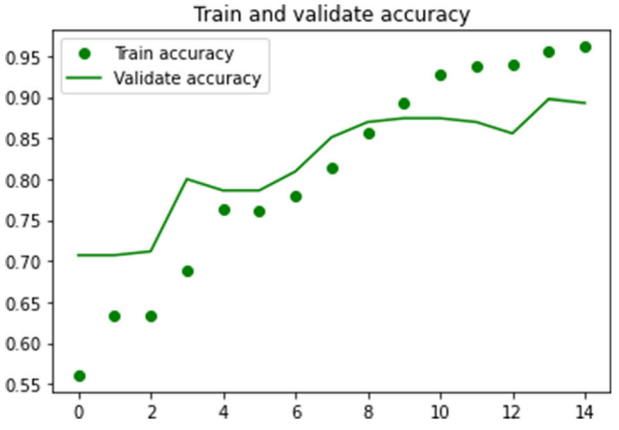
Following the N-gram analysis, the experimental method based on seven baseline methods and our proposed model is evaluated. The prime purpose of our research is to showcase evolving work on automatic recognition of suicidal content. In our baseline, we used single handcrafted features, such as TF-IDF, Bag of Words, Statistical Features and their combinations were applied on DT, RF, GB, Ensemble, ADABoost + DT, XGBoost and ADABoost + RF models. The major aim of amalgamating different NLP techniques is to investigate which features best suits the performance accuracy for suicide ideation. After that, the word2vec technique was applied to the LSTM, CNN or their combined models and their performances were evaluated. Figures 5 and 6 show the training and loss curves obtained for CNN - LSTM Model, 2- layer LSTM Model and Stacked CNN - 2 Layer LSTM Model with Word2Vec word embeddings respectively.

Table 3 shows the results of the classification models on suicide ideation identification tasks while Figs. 7, 8 and 9, and shows the graphical representation of the same. The table is divided into three categorisations depicting Machine Learning (TF-IDF + N-gram (1,2)) Classification Results pictographically represented in Fig. 7, Machine Learning (TF-IDF + N-gram (1,3)) Classification Results pictographically represented in Fig. 8 and Deep Learning (Word2Vec) Classification Results pictographically represented in Fig. 9. Each categorized corpus accommodates the values of accuracy, F-measure, recall and precision. It can be observed that the score of XGBoost is higher in the baseline as compared to other traditional text classification methods by taking into consideration both single and combined features.

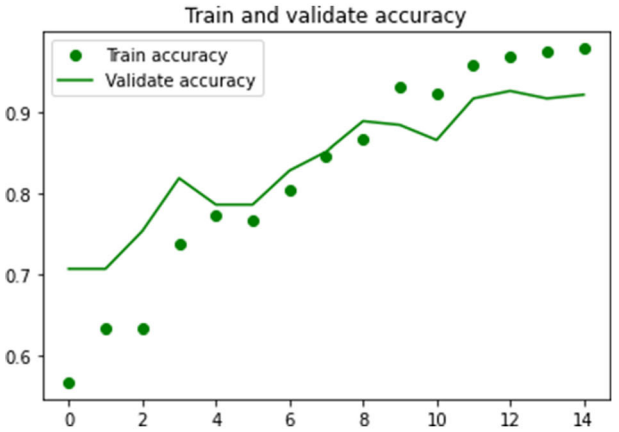
By performance assessment of machine learning methods in the baseline we noticed that the performance of XGBoost was better in terms of scoring more than various text classification techniques along with both single and combined features without the statistics. Accuracy achieved with XGBoost is 89.59% and 89.01% respectively in machine learning methods with F1-Scores as 89.53% and 89.27%.

According to this study, it is preferable to conduct a coarse line search over a single region in order to find the most appropriate size for the dataset for the models' parameters enhancement. Furthermore, examining the ReLU activation function, it gives a superior performance as well as the generalization, which was likewise applied in the research of [10, 52]. Max-pooling gives a better performance than other possible approaches for the classification tasks as indicated by our pooling approach.

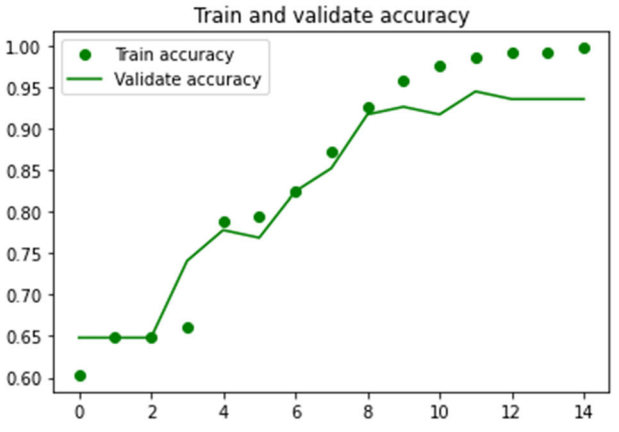
Contrasting the proposed model with the baseline, it can be concluded that the best classification result is attained with the Stacked CNN - 2 Layer LSTM combined model using Word2Vec as word embedding. The accuracy achieved in this case is 93.92% along with the values of F1-Score, Recall and Precision as 93.27%, 93.21% and 93.43% respectively. Further, it can be observed that the accuracy obtained with the CNN - LSTM combined model using word embedding is 89.30% which is clearly lesser than the accuracy obtained with the Stacked CNN - 2 Layer LSTM model. Results depict that the proposed model performs finer when compared to single CNN and LSTM Classifiers which have accuracies equal to 85.02% and 87.96% respectively. The reason for the enhanced accuracy in our proposed model is the use of stacked 2-Layer LSTM along with CNN. The results obtained also show that the combined model works better than the individual models. Our proposed model combines the strengths of both CNN and LSTM techniques and is also able to overcome the limitations such as overfitting and underfitting of the stated individual models which can also be a reason for the outstanding performance of the model. Hence, instead of using a single one of the deep



(a)

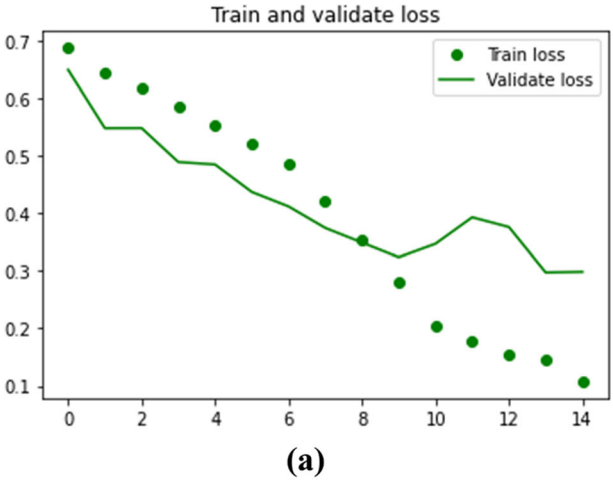


(b)

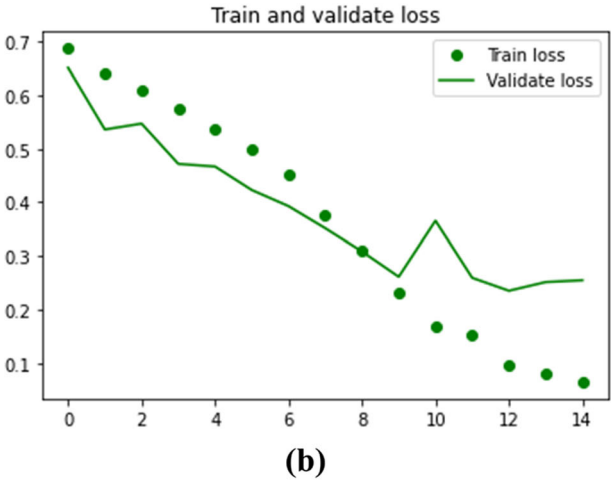


(c)

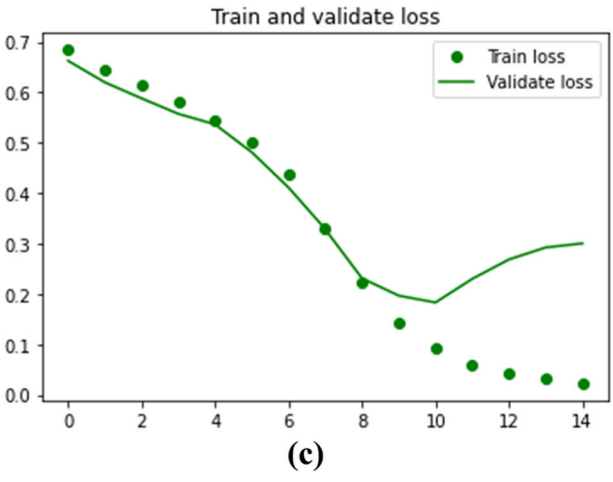
Fig. 5 **a-c.** Training Accuracy (●) and Validation Accuracy (——) vs Number of Epochs of CNN - LSTM Model, 2 - layer LSTM Model and Stacked CNN - 2 Layer LSTM Model with Word2Vec word embeddings respectively



(a)



(b)



(c)

Fig. 6 **a-c**. Training Loss (●) and Validation Loss (——) vs Number of Epochs of CNN - LSTM Model, 2 - layer LSTM Model and Stacked CNN - 2 Layer LSTM Model with Word2Vec word embeddings respectively

Table 3 Performance results of the Classification Models

CLASSIFICATION RESULTS (in %)					
Feature Type	Methods	Accuracy	F1-Score	Recall	Precision
Machine Learning (TF-IDF+N-gram (1,2))	DT	82.54	79.79	67.97	96.59
	RF	83.30	83.52	83.50	83.55
	GB	89.32	89.05	85.67	91.43
	Ensemble	82.75	82.94	82.66	83.21
	AdaBoost+DT	86.58	86.70	86.26	87.15
	XGBoost	89.59	89.53	87.70	92.72
	AdaBoost+RF	88.26	87.51	81.10	95.01
Machine Learning (TF-IDF+N-gram (1,3))	DT	82.20	79.38	67.54	96.24
	RF	82.35	82.53	82.18	82.88
	GB	88.74	88.49	85.30	90.10
	Ensemble	82.51	82.75	82.72	82.77
	AdaBoost+DT	85.36	85.10	82.42	87.96
	XGBoost	89.01	89.27	88.46	91.92
	AdaBoost+RF	88.95	88.32	81.88	95.86
Deep Learning (Word2Vec)	CNN	85.02	84.83	84.5	84.89
	LSTM	87.96	87.75	87.9	87.92
	2-Layer LSTM	92.09	91.91	91.87	92.01
	CNN-LSTM	89.30	89.2	89.14	89.22
	CNN+2 Layer -LSTM	93.92	93.27	93.21	93.43

DT = Decision Trees, RF = Random Forest, GB = Gradient Boosting, AdaBoost = Adaptive boosting, XGBoost = Extreme Gradient Boosting, LSTM = Long Short-term Memory, CNN = Convolutional Neural Networks

learning architectures, CNN and LSTM models [17] are combined to rectify the problems so as to improvise the stability, accuracy and also the predictive power of the proposed model. Using a stacked 2-Layer LSTM along with a preliminary CNN layer in our proposed model has further helped in improving the accuracy of our model which is one of the major reasons for using it instead of single layer.

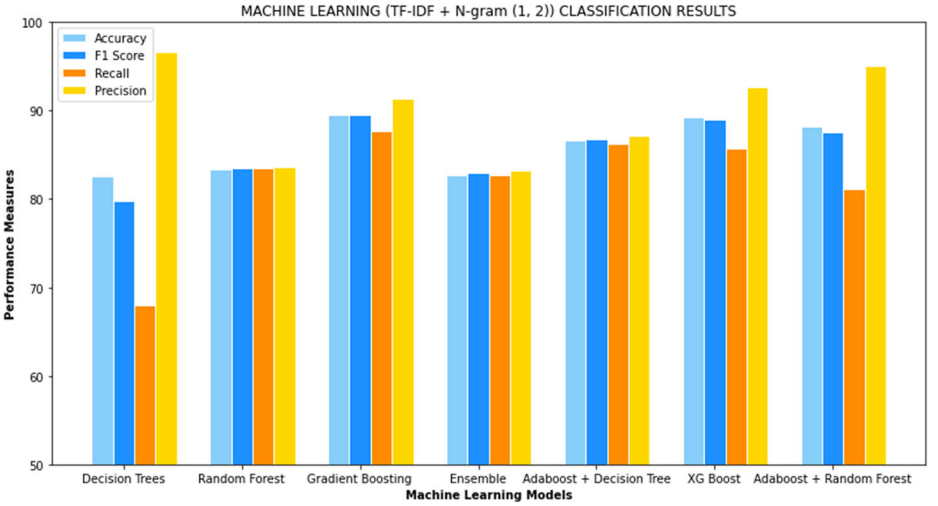


Fig. 7 Machine Learning (TF-IDF + N-gram (1,2)) Classification Results

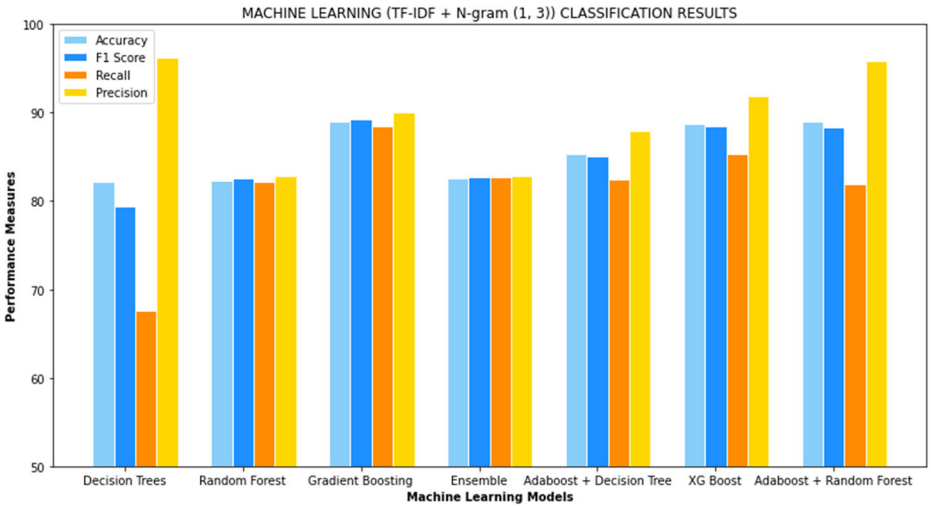


Fig. 8 Machine Learning (TF-IDF + N-gram (1,3)) Classification Results

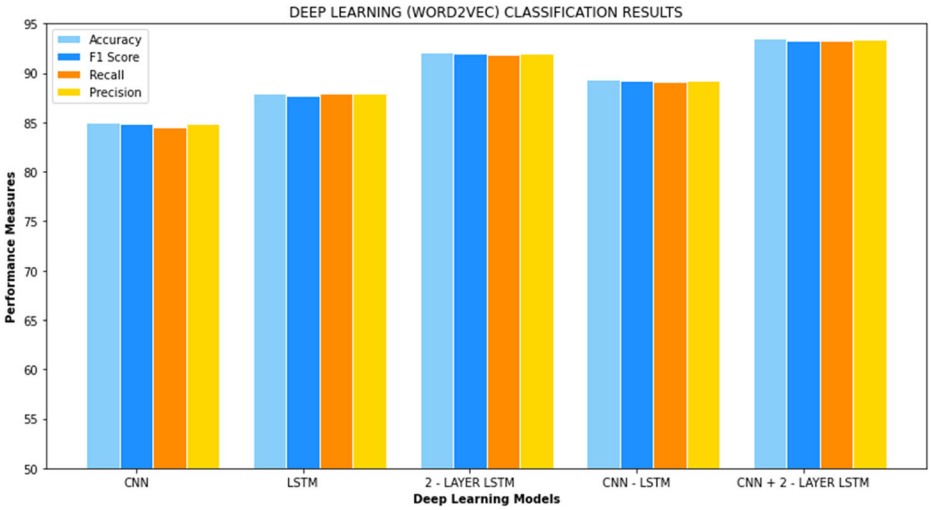


Fig. 9 Deep Learning (Word2Vec)) Classification Results

7 Conclusion and future scope

Our study aims to detect suicidal behaviour from the Twitter's tweet dataset. The incorporation of deep learning methods provides new directions for the enhancement of identification of suicidal behaviour and the likelihood for prior suicide avoidance. We proposed an approach to identify the presence of suicide ideation signs in tweets collection and concentrated on recognizing the most efficacious performance enhancement solutions. Our study takes part in this voyage towards the technological enhancement in convolutional linguistics to be shared within the research community and successfully implemented in mental health care. In this, different data representation approaches were used in order to transform the text of

tweets in such a way that our system can identify. Particularly, a closer link between suicidal thinking and language usage were specified by applying several text classification techniques. The experiment was described with LSTM and CNN networks set up on the top of word2vec features.

Based on the experimental results, the proposed Stacked CNN - 2 Layer LSTM hybrid model substantially enhances the accuracy of text classification. Combining the strengths of both LSTM and CNN algorithms and making up for their limitations, is the main reason that the model outperformed other machine learning classifiers. Foremost, it uses the CNN layer to extricate the local pattern and is able to undertake the text examining not only single words but also their amalgamations of various predefined sizes trying to learn their best fusion and expositions. Second, it takes advantage of the LSTM to preserve context details in a long text and fixes the problem of vanishing gradient. Using this perspective, we can corroborate that the hybrid model can successfully enhance the prediction results as we try to evince in our experiment.

In future work, we will implement the feature selection technique for selecting the best features in classifying and predicting suicide attempters with non-suicide attempters. Also, we will work on new datasets as well with correlated topics. We may also try to analyse the association between suicidal behaviour, family environment, etc. Additionally, different datasets will be utilized for further research with the ensemble of various deep learning classifiers and machine learning models like BERT for suicidal ideation classification along with different parameter optimization.

Declarations

Conflict of interest The Authors declare that there is no conflict of interest.

References

1. Ahmad S, Asghar MZ, Alotaibi FM, Awan I (2019) Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-centric Comput Inform Sci 9(1):24
2. Basu T, Murthy CA (2012) A feature selection method for improved document classification. In international conference on advanced data mining and applications (pp. 296-305). Springer, Berlin, Heidelberg
3. Beck AT, Kovacs M, Weissman A (1975) Hopelessness and suicidal behavior: an overview. Jama 234(11): 1146–1149
4. Bhat, H. S., & Goldman-Mellor, S. J. (2017) Predicting adolescent suicide attempts with neural networks. arXiv preprint arXiv:1711.10057
5. Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."
6. Chadha A, Kaushik B (2019) "A survey on prediction of suicidal ideation using machine and ensemble learning." Comput J
7. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
8. Choi SB, Lee W, Yoon JH, Won JU, Kim DW (2018) Ten-year prediction of suicide death using cox regression and machine learning in a nationwide retrospective cohort study in South Korea. J Affect Disord 231:8–14
9. Collobert R, Weston J (2008) A unified architecture for natural language processing: deep neural networks with multitask learning. In proceedings of the 25th international conference on machine learning (pp. 160-167)
10. Dahl GE, Sainath TN, Hinton GE (2013) Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8609-8613). IEEE

11. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M (2016) Discovering shifts to suicidal ideation from mental health content in social media. In proceedings of the 2016 CHI conference on human factors in computing systems (pp. 2098-2110)
12. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, Xu H (2018) Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Informa Decis Making 18(2):77–87
13. Freund Y, Schapire R, Abe N (1999) A short introduction to boosting. J-Japan Soc Artificial Intel 14(771–780):1612
14. Gehrmann S, Dernoncourt F, Li Y, Carlson ET, Wu JT, Welt J, Foote J, Moseley ET, Grant DW, Tyler PD, Celi LA (2018) Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. PLoS One 13(2):e0192360
15. Github. https://github.com/laxmimerit/twitter-suicidal-intention-dataset
16. Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) Deep learning (Vol. 1, no. 2). MIT press, Cambridge
17. Gupta S, Dinesh DA (2017) Resource usage prediction of cloud workloads using deep bidirectional long short term memory networks. In 2017 IEEE international conference on advanced networks and telecommunications systems (ANTS) (pp. 1-6). IEEE
18. He H, Lin J (2016) Pairwise word interaction modeling with deep neural networks for semantic similarity measurement. In proceedings of the 2016 conference of the north American chapter of the Association for Computational Linguistics: human language technologies (pp. 937-948)
19. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
20. Ji S, Yu CP, Fung SF, Pan S, Long G (2018) Supervised learning for suicidal ideation detection in online user content Complexity, 2018
21. Ji S, Long G, Pan S, Zhu T, Jiang J, Wang S (2019) Detecting suicidal ideation with data protection in online communities. In international conference on database systems for advanced applications (pp. 225-229). Springer, Cham
22. Kalchbrenner N, Grefenstette E, Blunsom P (2014) A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188
23. Klonsky ED, May AM (2014) Differentiating suicide attempters from suicide ideators: a critical frontier for suicidology research. Suicide Life Threat Behav 44(1):1–5
24. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
25. Li J, Zhang S, Zhang Y, Lin H, Wang J (2021) Multifeature fusion attention network for suicide risk assessment based on social media: algorithm development and validation. JMIR Med Inform 9(7):e28227
26. Lin G-M, Nagamine M, Yang S-N, Tai Y-M, Lin C, Sato H (2020) Machine learning based suicide ideation prediction for military personnel. IEEE J Biomed HealthIinform 24(7):1907–1916
27. Marks, M. (2019). Artificial intelligence based suicide prediction
28. Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S (2010) Recurrent neural network based language model. In Eleventh annual conference of the international speech communication association
29. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546
30. Morales M, Dey P, Theisen T, Belitz D, Chernova N (2019) An investigation of deep learning systems for suicide risk assessment. In proceedings of the sixth workshop on computational linguistics and clinical psychology (pp. 177-181)
31. Munikar M, Shakya S, Shrestha A (2019) Fine-grained sentiment classification using BERT. In 2019 artificial intelligence for transforming business and society (AITB) (Vol. 1, pp. 1-5). IEEE
32. Nordin N, Zainol Z, Mohd Noor MH, Lai Fong C (2021) A comparative study of machine learning techniques for suicide attempts predictive model. Health Inform J 27(1):1460458221989395
33. Onan A (2019) Topic-enriched word embeddings for sarcasm identification. In computer science on-line conference (pp. 293-304). Springer, Cham
34. Pestian J, Matykiewicz P, Grupp-Phelan J, Lavanier SA, Combs J, Kowatch R (2008) Using natural language processing to classify suicide notes. In proceedings of the workshop on current trends in biomedical natural language processing (pp. 96-97)
35. Pompili M, Innamorati M, Di Vittorio C, Sher L, Girardi P, Amore M (2014) Sociodemographic and clinical differences between suicide ideators and attempters: a study of mood disordered patients 50 years and older. Suicide Life Threat Behav 44(1):34–45
36. Rajesh Kumar E, Rama Rao KVSN, Nayak SR, Chandra R (2020) Suicidal ideation prediction in twitter data using machine learning techniques. J Interdiscip Mathema 23(1):117–125
37. Sawhney R, Manchanda P, Mathur P, Shah R, Singh R (2018) Exploring and learning suicidal ideation connotations on social media with deep learning. In proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis (pp. 167-175)

38. Sawhney R, Manchanda P, Singh R, Aggarwal S (2018) A computational approach to feature extraction for identification of suicidal ideation in tweets. In proceedings of ACL 2018, student research workshop (pp. 91-98)
39. Schapire RE, Singer Y, Singhal A (1998). Boosting and Rocchio applied to text filtering. In proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval (pp. 215-223)
40. Silver MA, Bohnert M, Beck AT, Marcus D (1971) Relation of depression of attempted suicide and seriousness of intent. Arch Gen Psychiatry 25(6):573–576
41. Sinha PP, Mishra R, Sawhney R, Mahata D, Shah RR, Liu H (2019) # Suicidal-a multipronged approach to identify and explore suicidal ideation in twitter. In proceedings of the 28th ACM international conference on information and knowledge management (pp. 941-950)
42. Sosa PM (2017) Twitter sentiment analysis using combined LSTM-CNN models. Eprint Arxiv, 1-9
43. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In proceedings of the 2019 conference of the north American chapter of the Association for Computational Linguistics: human language technologies, volume 1 (Long and short papers) (pp. 380-385)
44. Tadesse MM, Lin H, Xu B, Yang L (2020) Detection of suicide ideation in social media forums using deep learning. Algorithms 13(1):7
45. Wang Z, Qian X (2008) Text categorization based on LDA and SVM. In 2008 international conference on computer science and software engineering (Vol. 1, pp. 674-677). IEEE
46. Weng JC, Lin TY, Tsai YH, Cheok MT, Chang YPE, Chen VCH (2020) An autoencoder and machine learning model to predict suicidal ideation with brain structural imaging. J Clin Med 9(3):658
47. World Health Organization. (2018) National suicide prevention strategies: Progress, examples and indicators
48. World Health Organization (2018) National suicide prevention strategies: Progress, examples and indicators
49. Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853
50. Yang Y, Zheng L, Zhang J, Cui Q, Li Z, Yu PS (2018) TI-CNN: convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749
51. Yin W, Schütze H (2016) Multichannel variable-size convolution for sentence classification. arXiv preprint arXiv:1603.04513
52. Zeiler MD., Ranzato M, Monga R, Mao M, Yang K, Le QV, ..., Hinton GE (2013 On rectified linear units for speech processing. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 3517–3521). IEEE
53. Zhang, Y, Wallace B (2015) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820
54. Zhang X, Zhao J, Lecun Y (2015) Character-level convolutional networks for text classification. Adv Neural Inf Proces Syst 2015:649–657
55. Zhang, J.; Li, Y.; Tian, J.; Li T (2018) LSTM-CNN Hybrid Model for Text Classification. In Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China: pp. 1675
56. Zhu H, Xia X, Yao J, Fan H, Wang Q, Gao Q (2020) Comparisons of different classification algorithms while using text mining to screen psychiatric inpatients with suicidal behaviors. J Psychiatr Res 124:123–130

Affiliations

Bhavini Priyamvada[1] · Shruti Singhal[1] · Anand Nayyar[2,3] · Rachna Jain[4] · Priya Goel[1] · Mehar Rani[1] · Muskan Srivastava[1]

Bhavini Priyamvada
bhavinipriyamvada8@gmail.com

Shruti Singhal
shruti.singhal.2608@gmail.com

Rachna Jain
rachnajain@bpitindia.com

Priya Goel
priyagoel99@gmail.com

Mehar Rani
mehar212228@gmail.com

Muskan Srivastava
muskan.srivastava1904@gmail.com

[1] Computer Science Department, Bharati Vidyapeeth's College of Engineering, New Delhi, India

[2] Graduate School, Duy Tan University, Da Nang 550000, Viet Nam

[3] Faculty of Information Technology, Duy Tan University, Da Nang 550000, Viet Nam

[4] Information Technology Department, Bhagwan Parshuram Institute of Technology, New Delhi, India