A Study on Various Techniques Involved in Mental and Suicide Detection System: A Comprehensive Review

Shabir Ahmad Magray¹, Baijnath Kaushik²
School of Computer Science and Engineering, Shri Mata Vaishno Devi University, Jammu & Kashmir, India

Abstract:
Mental health is not limited to a particular country but it is a global issue, mental disorders may transform to suicidal ideation without successful therapy. To save the lives of people, early detection of mental disorder and suicidal ideation should be tackled. The prediction can be done using the clinical interviews between the experts and the targeted people and apply various machine or deep learning techniques for automatically detection of suicide risks in social content. In this paper, we will present the survey of methods for suicide detections. Various publicly available datasets will be also summarized in this paper.

Keywords: Suicide ideation, Deep learning, machine learning, social content.

1. INTRODUCTION
Mental health is not limited to a particular country, it is a global issue and mental health problems are increasingly concerned, making it second most cause of death in whole world. According to World Health Organization, to some extent, 1 in 4 people worldwide have mental disorders [1]. Without successful treatment, mental disorders could lead to suicide or suicide attempt. 9,00,000 individuals commit suicide every year across the world making suicide the most common causal agent of death in teenagers. Due to serious mental disorders, Number of suicide victims having mental disorder is 46% as per the US National Alliance on Mental illness [2]. Suicidal ideation determines whether a person has suicidal thoughts by analyzing the textual content written by him. One of the traditional way of assessing suicide risk is clinical interviews based consultation and psychological intervention. In order to detect the suicidal ideation, many clinical studies have been done and responses are classified [3]. This method doesn’t accurately predict the future suicidal behavior [4]. The failure of this is that suicidal individuals are not always willing to conceal their suicidal thoughts or intentions. Early discovery of self destruction ideation can distinguish individuals with self-destructive contemplations. Explanations behind the self destruction are perplexing and connected to a convoluted cooperation of numerous factors contemplations [5]. Due to the wide spread of internet activities and the use of social media, people start interacting to each other. People start expressing their feelings, emotions, sufferings and suicidal intentions. The online platform has become most helpful tool for people with mental health problems to provide support and feedback. By incorporating the techniques of machine learning and Artificial intelligence on social media data, suicide ideation can be predicted [6].

Detection on social media data is based on feature engineering [7], [8], sentiment analysis [9], [10] and deep learning [11], [12]. These methods need to extract the features from the available data and build appropriate models to correctly predict the mental health risks. The better the features selected the better will be the prediction made by the model. Artificial Intelligence is emerging technique and is being used to solve various computations problems. The suicide ideation problem can be addressed by using the Artificial Intelligence technique to understand the language used by the online users. The data used can be textual or tabular data and use Natural language Processing to represent the data. This comes with the drawback as the data available for training and testing is very limited. Artificial intelligence models some of the time learn statistical signs, but neglect the individuus’s real intentions. This paper provides a review of various categorization and methods used in suicide ideation. Various machine and deep learning methods along with the database used with various advantages and disadvantages have been summarized in Table

<table>
<thead>
<tr>
<th>Reference and year</th>
<th>Objective of the paper</th>
<th>Language /size of the dataset</th>
<th>Feature extraction methods</th>
<th>Classification method</th>
<th>Implementation results</th>
<th>Advantages/Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verena Venek et al [16]</td>
<td>Classify Suicide risk using clinical Patients interview</td>
<td>60 Audio recording of patient interviews. 30 suicidal and 30 non-suicidal</td>
<td>Feature extracted on transcript using LIWC. The features were of 80 categories. For voice signals COVAEP was used.</td>
<td>SVM, adaboost.</td>
<td>Overall accuracy was 73.3%</td>
<td><strong>Advantages</strong>&lt;br&gt;This method will help the experts to predict the mental state of patients.&lt;br&gt;<strong>Disadvantages</strong>&lt;br&gt;Individuals are not willing for a questionarrie.</td>
</tr>
<tr>
<td>Reference and year</td>
<td>Objective of the paper</td>
<td>Language /size of the dataset</td>
<td>Feature extraction methods</td>
<td>Classification method</td>
<td>Implementation results</td>
<td>Advantages/Disadvantages</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Stefan Scherer et al [17]</td>
<td>Classify suicidal and non suicidal adolescents using interview, out of which they used only data of 16 adolescents, 8 male and 8 female.</td>
<td>60 patients were interviewed. Out of which 30 were suicide patients and 30 were non suicidal.</td>
<td>Trim longer speechs to 30 second without interruption. Analyzed the jitter and glottal flow analysis.</td>
<td>HMM SVM</td>
<td>Accuracy of HMM 69% Accuracy of SVM 61%</td>
<td>Advantages: They got promising results with just the small set of extracted features. Disadvantages: A subset of the dataset was used instead of full dataset.</td>
</tr>
<tr>
<td>Daud Sikander et al [18]</td>
<td>Classify the suicidal and non suicidal patients using heart beat measureme nts.</td>
<td>The data was collected at Newcastle Mental Hospital by an expert. Dataset consists of 15 suicidal and non-suicidal patients.</td>
<td>Features extracted are: 1. Mean of heart rate 2. Standard deviation between two consecutive heart beats.</td>
<td>SVM, KNN</td>
<td>Accuracy of SVM is 80%. Accuracy of KNN is 71%</td>
<td>Advantages: Heart beat monitoring can be used with the clinical interviews to correctly classify the suicidal and non-suicidal patients.</td>
</tr>
<tr>
<td>Shaoxiong Ji et al [14]</td>
<td>Classify the online user data with supervised learning.</td>
<td>Data collected from online forums: reddit and twitter. The reddit data has 3549 suicidal ideation samples and large number of non ideation samples. Twitter data has 10,288 tweets containing 594 suicidal ideation tweets.</td>
<td>Word counts, POS, LIWC, TF-IDF, word2vec.</td>
<td>SVM, Random Forest, GDBT, XGBoost, Multi Layer feedforward network (MLFFN), LSTM.</td>
<td>On Reddit data set: SVM accuracy: 0.9132 Random Forest accuracy: 0.9357 GDBT accuracy: 0.9461 XGBoost accuracy: 0.9571 MLFFN accuracy: 0.9283 LSTM: 0.9266 On Twitter dataset: SVM accuracy: 0.9485 Random forest accuracy: 0.9638 GBDT accuracy: 0.9500 XGBoost accuracy: 0.9591 MLFFN accuracy: 0.9412 LSTM accuracy: 0.9108</td>
<td>Advantages: They showed that with more and more complicated features selection accuracy can be increased.</td>
</tr>
<tr>
<td>Reference and year</td>
<td>Objective of the paper</td>
<td>Language / size of the dataset</td>
<td>Feature extraction methods</td>
<td>Classification method</td>
<td>Implementation results</td>
<td>Advantages/Disadvantages</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------</td>
<td>--------------------------------</td>
<td>---------------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>X. Zhao et al [20]</td>
<td>Classify micro blog data using Double Input CNN.</td>
<td>Micro blog data.</td>
<td>Semantic and external features.</td>
<td>SVM,CNN , D-CNN</td>
<td>SVM accuracy=0.79 CNN accuracy =0.62 D-CNN accuracy=0.82</td>
<td>Advantage: It takes into account both semantic as well as external features. Disadvantage: It doesnot use all the external features.</td>
</tr>
<tr>
<td>H.C Shing et al[8]</td>
<td>Classify the online reddit data using CNN.</td>
<td>Training data of 625 users and testing data of 245 users</td>
<td>Word embedding</td>
<td>SVM and CNN</td>
<td>SVM F1 score=0.46 CNN F1 Score =0.42</td>
<td>Advantage: It uses CNN with max pooling strategy. Disadvantage: The dataset was restricted to reddit only.</td>
</tr>
<tr>
<td>Victor Ruiza et al[21]</td>
<td>Predicting the suicide risks of users from their reddit posts.</td>
<td>Reddit dataset containing the training data of 496 users (31,533) posts and test data of 125 users (9610) posts.</td>
<td>Clinical findings,word emotion association , Doc2</td>
<td>SVM, Naïve bayes,Gradient boosting , Random Forest CNN ,LSTM.</td>
<td>Macro F1 score of classifiers are: Naïve Bayes:0.422 Gradient boosting:0.412 Random Forest:0.395 SVM:0.432 CNN:0.367 LSTM:0.147</td>
<td>Advantage: NB turns out to be the best classifier under binary classification. Disadvantage: Both CNN and LSTM did not perform well as</td>
</tr>
</tbody>
</table>
Thus various methods were used in literature to classify the suicide ideation risk in individuals so that early intervention would help the experts to save their lives. Suicide ideation is a multi-class text classification problem involving the classes of none, low, moderate, and high. To classify the suicide risks, features have to be selected or use artificial neural network to learn the rich representation of the language used by the suicidal individuals. Thus by selecting the more useful features from the dataset and develop machine and deep learning models, to comprehend the language with self-destructive ideation better.

Suicidal ideation is acquiring fame and interest among researchers but still there are certain limitations:
1. Data deficiency is one of most critical issue in the current research. Most of the models use the supervised learning where data need to be labelled. There is very small labelled data available.
2. Among the data available, very small part of it contains the suicidal posts. Thus need is to have the balanced data contains both suicidal and non-suicidal in equal ratio.
3. Statistical method fail to learn the language as sometimes neglecting the certain important features.

Hence classifying the suicidal posts is a challenging process and more research need to be done in continuation of current advancement.

In this section, various studies were comprehensively compared to understand the prior art in this field. Various databases for suicidal ideation were also cited in Table 1. Next section will lists the methods used in the field and discusses them comprehensively.

2. RELATED WORK

Feature extraction and selection is the most important part of any text classification. Before feature selection, preprocessing of the data must be done. Suicide ideation can be viewed as an task of binary or multi-class classification.

2.1 Datasets:
The popular datasets used for suicide detection are mostly taken from social media. Few of them are as under:

2.2.1. Reddit Datasets: Reddit is a discussion forum and social news aggregator where people post in the form of posts, images and links. There are various topic categories and a particular category of topic is called subreddit. The most common subreddit known as “SuicideWatch” is extensively used for collection of suicide data. Shaoxiong Ji et al [14] released a dataset which contains a 3549 posts. H.C. Shing et al [8] released a University of Maryland Reddit dataset. The version 2 of the dataset was released in 2019, containing data of 11,129 users with 1,556,194 posts. Gaur et al [24] published a dataset of 500 redditors out of 2181 redditors.

2.2.2. Twitter Datasets: Twitter is a mainstream informal communication service, where numerous users talk about their suicide ideation. Ji et al [14] collected twitter dataset of 594 users out of 10,288 tweets. Coppersmith et al [25] collected the
data of 3200 tweets using twitter API. Vioules et al [26] collected a twitter dataset using twitter API which contains 5446 out of which 2381 are from distressed users and 3065 are from normal users.

2.2.3. ReachOut: Reachout is a platform which helps young adults with their mental health problems. Milne et al [27] published the Reachout dataset in CLPsych17 shared task, containing a training dataset of 65,756 discussion posts and a test dataset of 92,207 discussion posts.

2.2 Preprocessing Techniques
Preprocessing is first and foremost step used in text classification. The main purpose of the preprocessing is the dimensionality reduction. The preprocessing includes tokenization, change the data to lower case, removal of stop words and stemming or lemmatization.

2.3. Feature extraction. Feature extraction is the process of finding important differentiating attributes from the data and then transform them in a feature set that is usable by the classifier. In this section we will talk about the various feature extraction methods.

2.3.1. Bag of Words: This is one of the easiest feature extraction technique. It describes the occurrences of words in a document without bothering about the order of words. This featuring technique is used with machine learning approaches. The problem associated with the bag of words technique is the words with higher frequency becomes dominant and thus ignoring the domain specific words.

2.3.2. TF-IDF: This represents term frequency – inverse document frequency. It is utilized to correct the issue related with Bag of words

• Term Frequency (TF) is the frequency of words in a given text document.
• Inverse Term Frequency (IDF) is the measure of how unique or frequent term is a given text document.

This measure helps to distinguish between frequent and infrequent terms. Hence, the idf of unique terms is high and that of frequent terms is low.

2.3.3. Word2vec: Word embedding is constructed by word 2vec. Word embedding is one of most popular representation of document lexicons. This technique was created by Tomas Mikolov at google in 2013. The input is huge text and the output is a vector of hundreds of dimensions. Every unique word is mapped to the corresponding vector in the space. The words with similar context have the same value in vector space. There are two algorithms for word2 vec: continuous bag of words (CBOW) and skip gram of words. In CBOW, the prediction of target word is done from the context. In skip gram, the prediction of context words is done from the target word. Both these architecture have advantages and disadvantages. CBOW is faster and is preferred for smaller corpus. Skip gram is slower but is preferred for larger corpus.

2.3.4. GloVe: It stands for Global vectors and is log-bilinear regression model. This is one of the most famous word embedding technique. Unlike Word2Vec which relies on local statistics, Glove make use of both local and global statistics. This is implemented using gensim and uses the idea to train only on the non-zero elements in the word co-occurrence matrix.

3. DISCUSSION
There are various feature extraction and classification methods available. Deep learning techniques have shown promising results in suicide ideation. Combining various deep learning methods have also shown that the suicide prediction can be increased. The various classification algorithms and their results have been shown in the Table 1. The accuracy of a particular algorithm depends upon the feature extraction methods and the size of the datasets. The twitter datasets are not publicly available due the privacy concerns of twitter. Table 2 shows the datasets available for suicide detetion along with the size of data sets and the source of dataset.

Table 2. Description of datasets available for suicide ideation.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Source</th>
<th>Datasize</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper smith et al[13]</td>
<td>Twitter</td>
<td>3,200</td>
<td>NA</td>
</tr>
<tr>
<td>Vioules et al [26]</td>
<td>Twitter</td>
<td>5,446</td>
<td>NA</td>
</tr>
<tr>
<td>Milne et al [27]</td>
<td>Reachout</td>
<td>65,756</td>
<td>NA</td>
</tr>
<tr>
<td>Michael Mesfin Tadesse[23]</td>
<td>Reddit</td>
<td>7,201</td>
<td>NA</td>
</tr>
</tbody>
</table>

The general solution of the suicide ideation problem can be summarized in the following figure 1. It includes preprocessing to be done on the data and then extract the features using different methods and finally apply classification methods to predict the suicide risks among users. The combination of various algorithms can also be used to enhance the prediction.
4. CONCLUSION
Suicide is the leading cause of death among the young individuals in the whole world. Early detection of suicide is important and can save lives. This survey gives the comprehensive study of various techniques in suicide detection. It started from clinical patient interviews and continued to the online social media data by using the various feature extraction techniques and different classification algorithms. Due to the widespread of internet, tracing users online with mental health problems can be effective and important in saving their lives. Thus social media can serve as a tool in future to enhance the suicide detection. The need is to develop the new methods and techniques to completely mimic the language used by the suicide individuals using automated systems, in the hope suicide can be prevented.

5. REFERENCES


[14]. Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long. Supervised learning for suicidal ideation detection in online user content. Complexity, 2018, 2018


