INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND TECHNOLOGY
ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

Exploring Sentiment Analysis in the Context of Suicidal Thoughts: Leveraging NLP for Mental Health Insights.

Sushant Das

Student, Master of Computer Applications, School of Science Studies, CMR University, Bangalore, India.

Abstract:
Suicide is a critical issue in modern society. To save lives, early detection and prevention of suicide attempts should be addressed. Clinical methods based on interactions between social workers or experts and target persons as well as machine learning techniques with feature engineering or deep learning for automatic identification based on online social content are currently used to detect suicidal ideation. This work is the first survey to thoroughly explain and discuss each strategy in these groups. Suicide notes, electronic health records, questionnaires, and online user content are some examples of the data sources used to examine domain-specific applications of suicidal ideation detection. To aid further research, several specific tasks and data sets are provided and summarized. Finally, we summarize the limitations of current work and provide an outlook of further research directions.

Keywords
Suicidal ideation, sentiment analysis, Suicidal thoughts, Depression, Mental health, Crisis hotline, Mental health awareness.

I. Introduction:

Mental health problems like depression and anxiety are causing growing worry in modern society because they are particularly severe in developed nations and emerging economies. Without adequate treatment, severe mental illnesses can lead to suicidal thoughts or even suicide attempts. Some online posts are quite depressing and lead to undesirable occurrences like cyberbullying and stalking. Since such poor information frequently engages in some type of social cruelty, the consequences can be serious and dangerous, resulting in gossip or even mental harm. Cyberbullying and suicide are related, according to research [1]. Overexposed victims who experience too many unfavorable incidents or messages may become hopeless and sad, and even worse, some may end their lives. There are many complex factors that contribute to suicide. Although depressed individuals are much more likely to kill themselves, many people without depression are also capable of having suicidal thoughts[2]. The American Foundation for Suicide Prevention (AFSP) classifies suicide-related causes into three groups: historical, environmental, and health-related factors [3]. According to Ferrari et al. [4], substance use disorders and mental health conditions are contributing factors to suicide. In their extensive analysis of the psychology

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

of suicide, O'Connor and Nock[5] categorized psychological hazards as personality and individual variations, cognitive variables, social factors, and unfavorable life events.

Suicidal Ideation Detection (SID) uses tabular information about a person or text that was authored by a person to identify whether they have suicidal ideation or thoughts. As social media and online anonymity have developed, an increasing number of individuals turn to interact with others on the Internet. People are increasingly expressing their emotions, pain, and suicidal thoughts through online communication platforms. As a result, social media content mining can help with suicide prevention and online channels have started to operate as a surveillance tool for suicidal ideation [6].

Strange social phenomena are starting to appear, such as internet communities agreeing on copycat suicide and self-mutilation. For instance, the "BlueWhaleGame"1 social network phenomenon from 2016 involved numerous tasks (such as self-harming) and ultimately resulted in suicide for game participants. The social problem of suicide claims hundreds of lives each year. Therefore, it's important to identify suicidality and stop suicide attempts before people take their own lives. The best means of averting prospective suicide attempts are seen to be early discovery and treatment.

Potential suicide victims may act out their thoughts of killing themselves in role-playing, fleeting thoughts, and suicide plots. Before a tragedy occurs, these risks of intentions or behaviors can be discovered through the identification of suicidal thoughts. In addition to highlighting the statistical limits of ideation as a screening tool, Mc Hughetal's meta-analysis[7] also noted that people's presentation of suicidal thoughts is a sign of their psychological distress. Early warning signs of suicide ideation can be effectively detected, allowing social workers to identify those who are thinking about harming themselves and establish a line of communication. The reasons for suicide are complicated and attributed to a complex interplay between many elements [5], [8]. Numerous researchers carried out psychological and clinical tests [9] and categorized the results of questionnaires [10] to identify suicidal ideation. Artificial intelligence (AI) and machine learning approaches can predict a person's chance of suicide based on their social media data[11], which can help us better understand people's motivations and open the door for early intervention.

Deep learning [16], [17], [18], sentiment analysis [14, 15], and feature engineering [12, 13] are the main techniques used for social content detection. In order to choose features or create artificial neural network topologies for learning rich representation, these techniques typically require heuristics. The current research trend focuses on identifying more beneficial elements from the health records of individuals and creating neural architectures to better grasp the language associated with suicidal ideation.

Mobile technologies have been researched and used to prevent suicide, such as the Black Dog Institute's mobile suicide intervention software iBobbly [19], which is only one example. Other suicide prevention technologies that are incorporated with social networking services have also

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

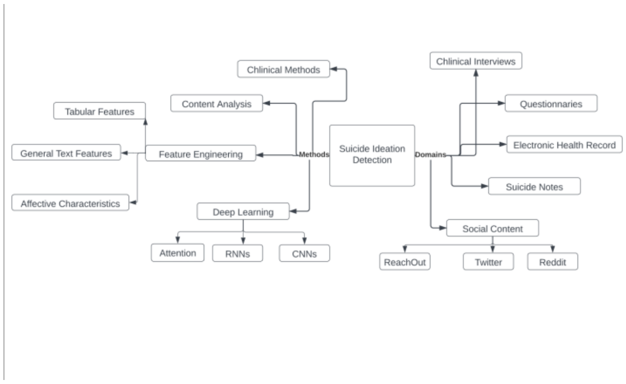
been created, such as Woebot 4 and Samaritans Radar 3. The former was a Twitter plugin that was subsequently removed due to privacy concerns. to watch over alarm posts. The latter is a Facebook chatbot designed to reduce people's anxiety and depression using cognitive behavioral therapy and natural language processing (NLP) methods It is inevitable that using cutting-edge AI technology to detect suicide ideation would raise privacy and ethical issues [20, 21]. Three moral concerns were brought up by Linthicum et al. [22], including the impact of prejudice on machine learning algorithms, the estimation of the timing of suicide acts, and the moral and legal issues brought up by false positive and false negative predictions. It is challenging to provide ethical answers for AI since these challenges call for algorithms to strike a balance between conflicting values, concerns, and interests[20]. Numerous difficult social issues have been solved with AI. One of the potential uses of artificial intelligence for the benefit of society is the detection of suicidal ideation, which needs to be addressed in order to meaningfully enhance people's well-being. Feature selection on tabular and text data as well as representation learning on natural language are among the research's issues. To categorize suicide risks, numerous AI-based techniques have been used. But there are still some difficulties. Benchmarks for suicidal ideation detection training and evaluation are scarce. Sometimes AI-powered programs pick up statistical cues but are unable to discern people's intentions. Additionally, a lot of brain models are difficult to comprehend. This study examines various ways to identify suicidal thoughts from the standpoint of AI, machine learning, and specific domain applications with social implications. Fig. 1 displays the categorization from these two angles.
This article offers a thorough analysis of the rapidly growing topic of machine learning-based suicidal ideation identification. It suggests an overview of recent scientific developments and a forecast for further study. The following is a summary of the results of our survey.

• To the best of our knowledge, this is the first study that examines suicidal ideation detection from a machine learning perspective, including its approaches and applications.

•We introduce and talk about traditional content analysis and contemporary machine learning methods, as well as how they are used to survey data, EHR data, suicide notes, and online social media content.

• We list the current and less-explored tasks and talk about their drawbacks. We also present an overview of the current data sets and a forecast for the future paths of this field's study.
The remainder of the study is divided into sections that introduce methodologies, summarize applications, list particular tasks and datasets, discuss our findings, and suggest some next possibilities.

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847



II. Related Work:

The intersection of sentiment analysis and suicidal thoughts has drawn significant attention from researchers, health professionals, and technology developers. In this section, we review the existing body of work in this domain, highlighting the methodologies, findings, and contributions made in the context of Sentiment Analysis and Suicidal Thoughts using Natural Language Processing (NLP).

Sentiment Analysis and Mental Health:
Studies exploring the connection between sentiment analysis and mental health have provided a foundational framework for understanding the emotional states of individuals who may be experiencing suicidal thoughts. De Choudhury et al. (2016) have notably demonstrated the feasibility of using sentiment analysis to predict depression risk and suicidal ideation. By analyzing social media data, linguistic markers, and sentiment patterns, the study achieved promising results in identifying users at risk of developing suicidal thoughts and depression.

Similarly, the work of Park et al. (2012) has explored sentiment analysis techniques for detecting signs of depression and suicidal thoughts within online communities. Their approach involved the analysis of sentiment in users' forum posts and comments. They found that individuals expressing negative sentiments, particularly those involving hopelessness and despair, were more likely to exhibit suicidal ideation. These studies underscore the potential of sentiment analysis in predicting and understanding mental health issues.

Social Media and Online Platforms:
Given the proliferation of online communication, social media platforms and online forums have become fertile grounds for studying sentiment and mental health. Coppersmith et al. (2018) conducted an extensive study involving millions of tweets to develop a large-scale lexicon of

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847
terms indicative of mental health issues, including suicidal thoughts. Their work not only provided valuable resources for researchers but also emphasized the use of public social media data to monitor emotional states at scale.

Additionally, the Crisis Text Line dataset, introduced by Smith et al. (2016), has emerged as a crucial resource in the study of sentiment analysis and suicide prevention. The dataset comprises anonymized crisis helpline text conversations, allowing researchers to apply NLP techniques to identify individuals in crisis and develop predictive models for suicide risk. This dataset has facilitated real-time interventions and a deeper understanding of emotional states in crisis contexts.

Ethical and Privacy Considerations:

Several studies have also delved into the ethical and privacy considerations inherent in sentiment analysis and mental health research. Guntuku et al. (2019) explored the trade-off between predictive accuracy and privacy preservation in the context of using social media data for mental health predictions. Their research highlighted the need for responsible data handling practices to protect the privacy of individuals while utilizing sentiment analysis for suicide prevention.

Cross-Cultural Perspectives:

While a majority of the research in this domain has been conducted within the context of English-language data, cross-cultural perspectives are essential. Forte et al. (2020) conducted a comparative study to understand cultural differences in expressions of mental health on social media. Their research emphasized the need to adapt sentiment analysis techniques to different linguistic and cultural contexts, especially when addressing sensitive topics like suicidal thoughts.

The existing body of work in Sentiment Analysis and Suicidal Thoughts using NLP provides valuable insights and methodologies for early detection and intervention. These studies underscore the potential for sentiment analysis to play a vital role in suicide prevention and mental health support. However, ethical considerations, privacy concerns, and cross-cultural perspectives must be taken into account in future research, emphasizing the importance of ongoing exploration and innovation in this critical domain.

III. Literature Review:

Sentiment Analysis and Suicidal Thoughts have gained significant attention in recent years as the potential for early detection and intervention in mental health issues through Natural Language Processing (NLP) techniques becomes increasingly evident. This literature review provides an overview of existing research, methodologies, findings, and gaps in the context of Sentiment Analysis and Suicidal Thoughts using NLP.

Sentiment Analysis and Suicidal Thoughts:

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

Sentiment analysis, a subfield of NLP, involves the automatic extraction of emotional information from textual data. In the realm of mental health, sentiment analysis plays a pivotal role in assessing individuals' emotional states, thereby offering the potential to identify signs of distress, particularly those related to suicidal thoughts. Accurate sentiment analysis can serve as a valuable tool for detecting and assisting individuals at risk of self-harm or suicide.

Suicidal Thoughts and Mental Health:

Suicidal thoughts are a concerning aspect of mental health, associated with various mental health disorders. According to the World Health Organization, suicide remains a global public health issue, claiming hundreds of thousands of lives annually, with countless more suicide attempts. Detecting and understanding suicidal thoughts is crucial for timely intervention and prevention.

Natural Language Processing (NLP):

NLP NLP is a subfield of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language. NLP provides the computational tools and methodologies necessary for processing and analyzing extensive volumes of textual data, enabling the extraction of valuable insights from diverse sources, including social media, online forums, chat logs, and clinical records.

Previous Research on Sentiment Analysis and Suicidal Thoughts:

Several studies have delved into the realm of sentiment analysis as it pertains to suicidal thoughts and mental health. For instance, De Choudhury et al. (2016) utilized data from social media platforms to predict depression risk and suicidal ideation through linguistic markers and sentiment analysis. The study found promising results in identifying users at risk of suicidal thoughts and depression.

Data Sources and Datasets:

Research in this domain often relies on publicly available data sources, such as social media platforms (e.g., Twitter, Facebook), online forums, crisis helplines, and anonymous chat applications. Datasets like the Crisis Text Line dataset and the Reddit Suicide Watch dataset have been commonly used for training and validating sentiment analysis models.

Methodologies and Techniques:

NLP techniques have been applied in various ways, including text preprocessing, sentiment polarity analysis, topic modeling, and machine learning algorithms. Sentiment lexicons like the

VADER (Valence Aware Dictionary and sentiment Reasoner) have been widely adopted to gauge emotional content in text data related to suicidal thoughts.

Sentiment Analysis and NLP hold immense potential in the context of mental health, particularly in the early detection of suicidal thoughts. Existing research has made significant strides in harnessing these technologies for mental health applications, but challenges and gaps underscore the ongoing importance of research and innovation in this critical domain. Researchers and practitioners must continue to work towards more accurate, reliable, and ethically sound methods for detecting and supporting individuals at risk of suicidal thoughts.

While existing research has produced promising results and insights, notable gaps persist in the field. The need for real-time intervention systems, larger and more diverse datasets, including clinical records and private communication channels, and research into cross-cultural differences in the expression of suicidal thoughts remain areas warranting further exploration.

IV. Proposed Method:

Due to an increase in suicide rates in recent years, suicide detection has attracted the interest of numerous experts and has received substantial research from a variety of angles. The research methods used to study suicide also cover a wide range of disciplines and methodologies, such as clinical approaches including patient-clinician contact [9] and automatic detection using user-generated information (mostly text) [12], [17]. Techniques from machine learning are frequently used for automatic detection.

Clinical techniques, such as face-to-face interviews and self-reports, are used in traditional suicide detection. Veneket et al. [9] created a five-item universal questionnaire for the assessment of suicidal risks and used the patients' responses to apply a hierarchical classifier to ascertain their suicidal intentions. Verbal and acoustic information can be used during face-to-face interactions. In order to distinguish between young people who were suicidal and those who were not, Scherer[23] looked at the prosodic speech patterns and voice quality in dyadic interviews.

Other clinical approaches identify functional magnetic resonance imaging-based neuronal representations of words associated with death and life [24], investigate resting state heart rate from converted sensory signals [24], and analyze event-related instigators from converted EEG signals [26]. Understanding the psychology of suicidal behavior is another part of therapeutic treatment[5], but it strongly depends on the clinician's expertise and face-to-face engagement. Clinical interviews combined with suicide risk assessment instruments can provide useful hints for suicide prediction[27].

Weibo, a Chinese version of Twitter, was used by Tan et al.[28] to conduct an interview and survey study to examine how suicide attempters responded to intervention through direct messages.

***A Content Analysis**:*

Social media posts by users give detailed information about them as well as their preferred languages. One can gain insight into language use and linguistic cues of suicide attempters by exploratory data analysis on user-generated content. In-depth analysis of posts relating to suicide involves topic modeling, statistical linguistic aspects, and lexicon-based filtering.

Keyword filtering [29], [30], and phrase filtering [31] are made possible by a manually created suicide-related keyword dictionary and lexicon. The words "kill," "suicide," "feel alone," "depressed," and "cutting myself" are all associated with suicide.Vioule'setal.Using a Twitter dataset with annotations, [3]built a point-wise mutual information symptom lexicon.

Gunn and Lester [32] examined tweets from the day before a suicide attempt passed away. Coppersmithetal. [33] examined data from the same platform's language usage. Strongly negative emotions, anxiety, and hopelessness, as well as other social factors like family and friends, may all play a role in suicidal ideation. Ji et al.'s [17] analysis of suicide-related content using topic modeling and word cloud visualization revealed that personal and social problems are discussed in relation to suicide. Colombo et al.'s [34] analysis of the Twitter social network's connectivity and communication graphs. An exploratory analysis of language usage and emotional content on Twitter was presented by Coppersmith et al. [35]. Other approaches and techniques include speech pattern analysis [39], social media content detection [36], reply bias assessment through linguistic cues [37], human-machine hybrid method for analysis of the language effect of social support on suicidal ideation risk [38], and Google Trends analysis for suicide risk monitoring [36].

The purpose of text-based suicide classification is to ascertain whether candidates have suicidal thoughts through their posts. NLP and machine learning techniques have both been used in this area.

Table 1 Features Tabular data for the detection of suicidal thoughts include survey responses and structured statistical data collected from websites. Such organized data can be utilized directly as characteristics for regression or classification. In order to categorize suicide and control groups based on user characteristics and social behavior variables, Masuda et al. [40] used logistic regression.

B Feature Engineering:

The purpose of text-based suicide classification is to ascertain whether candidates have suicidal thoughts through their posts. NLP and machine learning techniques have both been used in this area.

1.Tabular Features:

Responses to questionnaires and structured statistical data gathered from websites make up tabular data used to detect suicidal ideation. Direct use of these structured data as features for regression or classification is possible. Based on user characteristics and social behavior indicators, Masuda et al. [40] used logistic regression to identify the suicide and control groups.

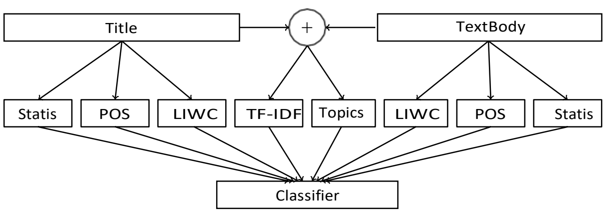
ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

The authors discovered that in a Japanese SNS, factors including homophily, local clustering coefficient, and community number have a more significant impact on suicidal ideation. In order to evaluate suicide risk factors, Chattopadhyay [41] employed the Pierce Suicidal Intent Scale (PSIS) and carried out regression analysis. Surveys are a reliable source of tabular features. Delgado-Gomezetal.[42] employed the Holmes Rahe Social Readjustment Rating Scale and the International Personality Disorder Examination Screening Questionnaire. According to Beck's suicide intent scale, Chattopadhyay[43] advocated using a multilayer feed-forward neural network, as depicted in Fig.2, to classify suicidal intention indications.

***2. General Text Features*:**

Extraction of features from unstructured text is another use of feature engineering. N-gram features, knowledge-based features, syntactic features, context features, and class-specific features make up the majority of the features.[44].Abbouteetal.[45]created a list of terms for nine suicidal subjects for vocabulary feature extraction. Okhapkinaetal. [46]created a dictionary of words with a suicidal theme.

Neural network with feature engineering

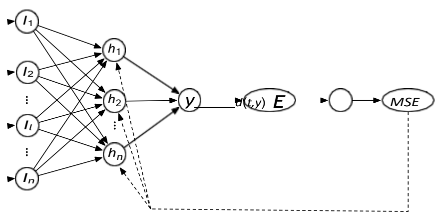


They invented singular valued e-composition (SVD) for matrices and term frequency-inverse document frequency (TF-IDF) matrices for communications. Mulholland and Quinn[47] built a classifier to predict the propensity of a lyricist's suicide by extracting language and syntactic factors. To detect cyber suicide in Chinese microblogs, Huang et al.[48] constructed a psychological lexicon dictionary by expanding How Net (a collection of common sense words).

To detect suicide in Sina Weibo, the topic model [49] is combined with various machine learning methods. Ji et al.'s [17] extraction of a number of informative sets of features, such as statistical, syntactic, linguistic inquiry and word count (LIWC), word embedding, and topic features, was used to create classifiers as depicted in Fig. 2b, which compares four conventional supervised classifiers. As part of their Bag of Words (BoWs) analysis, Shinget al. [13] collected a number of features, including empathy, readability, syntactic features, topic model posteriors, word embeddings, linguistic inquiry and word count, emotion features and lexicon of mental disorder.

SVM [44], artificial neural networks (ANN) [50], and conditional random field (CRF) [51] are models for detecting suicidal thoughts with feature engineering. Taietal [50] chose a number of characteristics, including a candidate's family history of mental illness, their history of self-harm

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847
and suicidal ideation, as well as their religious convictions and family status. The effectiveness of several multivariate approaches was compared by Pesti et al. [52] using word counts, POS, concepts, and readability ratings as performance indicators. The classification techniques of logistic regression, random forest, gradient boosting decision tree, and XG Boost were also compared by Jiet al. [17] in a similar manner. Braithwaite et al.Machine learning techniques that have been verified [53]can successfully identify high suicide risk.



Classifier with feature engineering

3. Affective Characteristics:
Researchers in both mental health and computer science have paid close attention to the fact that affective traits are among the most noticeable variations between those who attempt suicide and healthy people. Liakataetal. [51] employed manual emotion categories, such as anger, grief, hopefulness, happiness/peace, fear, pride, abuse, and forgiveness, to identify the emotions in suicide notes. In order to develop fine-grained sentiment analysis, Wang et al. [44] combined characteristics of factual (2 categories) and emotive aspects (13 categories). Similar to this, Pestian et al. [52] recognized abuse, anger, blame, fear, guilt, hopelessness, sorrow, forgiveness, happiness, peace, hopefulness, love, pride, thanksgiving, instructions, and information as feelings. In order to examine the emotional features that have accumulated in suicide blogs and to identify suicidal intentions from a blog stream, Ren et al. [14] suggested a complex emotion topic model. The eight basic emotions of joy, love, expectancy, surprise, worry, sorrow, rage, and hatred with a five-degree intensity were specifically examined by the writers, along with emotion accumulation, covariance, and transition.

C Deep Learning:

Deep learning has had tremendous success in a variety of fields, including computer vision, natural language processing, and medical diagnosis. It is a crucial technique for automatic suicidal ideation identification and suicide prevention in the field of suicide research. It does not require complex feature engineering approaches to successfully learn text characteristics automatically. In addition, some researchers feed extracted data into deep neural networks. For instance, Nobles et al. [54] supplied word occurrence and psycholinguistic variables into the multi-layer perceptron (MLP). Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT) are

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

examples of common deep neural networks (DNNs), as shown in Fig. 3a,3band 3c. Typically, prominent word embedding methods like word2vec [55] and GloVe [56] are used to embed natural language text into distributed vector space. Shingetal.[13]used user-level CNN to encrypt user posts using filters with sizes 3, 4, and 5. Textual sequences are encoded using a widely used RNN version called the long short-term memory (LSTM) network, which is subsequently processed through fully linked layers for classification[17]. New techniques for detecting suicidal thoughts integrate DNNs with other cutting-edge learning paradigms. Jietal [57] developed model aggregation techniques for updating CNNs and LSTMs, two types of neural networks with the aim of identifying suicidal ideation in private talking rooms. Decentralized training, on the other hand, is dependent on coordinators in chat rooms labeling user posts for supervised training, which can only be used in the most basic cases. Using unsupervised or semi-supervised learning techniques could be a preferable option. By predicting the gender of users as an auxiliary task, Benton et al. [16] predicted suicide attempts and mental health using neural models within the context of multi-task learning. In order to increase performance, Gauretal.[58]in combined external knowledge bases and suicide-related data onto a text representation. Coppersmithetal. developeda [59]. The deep learning model includes a self-attention mechanism to capture the most informative subsequence, bidirectional LSTM for sequence encoding, and GloVe for word embedding. Sawhney and others [60] LSTM, CNN, and RNN for suicidal ideation detection. Tadesseetal[61] also used the LSTM-CNN model. For encoding text and danger indicators, Jietal[62] presented an attentive relation network with LSTM and topic modeling.

Numerous well-known DNN architectures were used in the 2019 CL Psych Shared Task [63]. Pretraining was examined by Hevia et al. [64] employing a variety of models, including GRU-based RNN. Several well-known deep learning models, including CNN, LSTM, and Neural Network Synthesis (Neu Net S), were examined by Morales et al. [65]. Dual-context modeling utilizing hierarchically attentive RNN and BERT was proposed by Matero et al. [66]. The so-called hybrid method, which combines representation learning techniques with minor feature engineering, is another sub-direction. A hybrid categorization model of the behavioral model and the suicide language model was put forth by Chen et al. [67]. For categorizing individuals who attempt suicide and have depression, Zhao et al. [68] suggested the D-CNN model using word embedding and external tabular characteristics as inputs.

V. Experimental Setup:

A. Tasks:

A1. Suicide Text Classification: The first task, which entails binary and multi-class classification, can be seen as a domain-specific application of general text classification. While multi-class suicidality classification carries out a precise suicide risk assessment, binary suicidality classification simply assesses whether or not text contains suicidal ideation. For instance, some studies categorize suicide risk into no, low, moderate, and severe levels. As an alternative, it can take into account four different classifications based on mental and behavioral

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847
processes, namely non-suicidal, suicidal thoughts and wishes, suicidal intentions, and suicidal acts or plans. Risk assessment using information from postings with several aspects of suicide is another subtask. Gilatetal[94] manually tagged suicidal posts with many aspects labels, such as mental pain, cognitive attribution, and level of suicidal risk, according to the description of features of suicidal communications. Scaled into [0, 7], mental agony encompasses loss of control, intense loneliness, emptiness, narcissistic wounds, irreversible energy loss, and emotional flooding. If there is no evidence of attribution, cognitive attribution is the unmet demands related to interpersonal interactions.

A2. Reasoning Suicidal Messages: DNNs have been utilized by large-scale data mining and machine learning algorithms to reach remarkable results. Simple feature sets and classification algorithms, however, are not sufficiently foretelling to identify complex suicidal intentions. Suicidal communications must be reasoned using machine learning approaches in order to gain a deeper understanding of the innermost being and the causes that lead people to consider suicide. In order to better forecast suicide factors, this endeavor intends to use interpretable methodologies to explore suicidal aspects and combine common sense thinking. The specific tasks involve an automatic summary of the suicide risk, an explanation of the risk of suicide in mental suffering, and features of suicide cognition that are related to attribution.

A3. Suicide Attempter Detection: The two tasks described above concentrate on a single text. However, identifying suicide attempters is the main goal of suicidal ideation detection. As a result, it is crucial to achieve user-level detection, which entails two steps: graph-based user-level multi-instance suicidality identification and graph-based user-level suicide attempt detection. The former does multi-instance learning over a bag of messages using a bag of user postings as input. Afterward, it locates suicide attempters in a particular social graph created by users' social network interactions. It takes into account the connections between social users and can be seen as a different graph categorization issue.

A4. Generating Response: Suicide prevention and intervention are the ultimate goals of the detection of suicidal ideation. At midnight, a lot of people who intend to kill themselves usually post about their misery. To enable prompt social care and relieve their suicide purpose, it is also necessary to develop intelligent responses for counseling possible suicidal victims. Gilatetal [94] described eight different categories of response tactics, including emotional support, group support, empowerment, interpretation, inducing cognitive change, persuasion, advising, and referral. In order to accomplish this objective, machine learning techniques—in particular, sequence-to-sequence learning—must be able to adopt efficient response strategies that will lead to better responses and reduce suicidality among people. This response-creation technique can also create suggestions for social workers or volunteers to compose sensible responses when they return to the internet.

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

***A5. Mental disorders and Self-harmrisk**:* Suicidal ideation is closely related to the risks of self-harm and mental health problems. Therefore, spotting risks for serious mental illnesses or self-harm is equally crucial. These studies include self-harm detection[96], stress period and event detection[97], depression knowledge graph construction[98], depression and anxiety correlation analysis[99], and depression detection[95].

B. Datasets:

a. Reddit: Reddit is a publicly accessible internet community that collects social news and forum posts. Each area of interest within a topic is referred to as a subreddit, and there are numerous topic categories on it. "Suicide Watch" (SW)8 is a subreddit that is frequently used for additional annotation as positive samples. Suicide-free posts are pulled from other well-known subreddits. J etal. A dataset with 3,549 posts including suicide thoughts was published by [17]. With 11,129 users and 1,556,194 posts in total, Shing et al.'s [13] UMD Reddit Suicidality Dataset was published. They also selected 934 users for additional annotation. Aladag˘etal.Using Google Cloud Big Query, [100] gathered 508,398 posts and manually annotated 785 of them.

8https://reddit.com/r/SuicideWatch

b. Twitter: Twitter is a well-liked social networking site where many users also discuss having suicide thoughts. In terms of post length, anonymity, and the nature of communication and interaction, Twitter is very different from Reddit. Coppersmith et al. [33] have compiled data on Twitter users who have experienced depression and suicide thoughts. Out of a total of 10,288 tweets, Ji et al. [17] gathered an unbalanced sample of 594 tweets with suicide ideation. Utilizing the Twitter streaming API, Vioule'set al. gathered 5,446 tweets, of which 2,381 and 3,065 were sent by people in distress and other users, respectively. However, due to Twitter rules, the majority of Twitter-based datasets are no longer accessible.

c. ReachOut: ReachOut Forum9 is a peer support platform provided by an Australian mental health care organization. The CLPsych17 shared task was where the ReachOut dataset[101] was first made public. Participants were initially given a training dataset of 65,756 forum posts, of which 1188 were annotated manually with the expected category, and a test set of 92,207 forum posts, of which 400 were identified as requiring annotation. The specific four categories are described as follows. 1) crisis: the author or someone else is at risk of harm; 2) red: the post should be responded to as soon as possible; 3) amber: the post should be responded to at some point if the community does not rally strongly around It; 4) green: the post can be safely ignored or left for the community to address.

9https://au.reachout.com/forums

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

VI. Conclusion:

In today's society, there is still much that needs to be done to prevent suicide. A crucial and successful method of preventing suicide is the early identification of suicidal ideation. This survey examines the various techniques for detecting suicidal ideation from a wide range of angles, including clinical approaches like patient-physician interaction and medical signal sensing, textual content analysis techniques like lexicon-based filtering and word cloud visualization, feature engineering techniques like tabular, textual, and affective features, and deep learning-based representation learning techniques like CNN- and LSTM-based text encoders. There are four main applications introduced: questionnaires, electronic health records, suicide notes, and online user contributions.

The majority of this field's research has been done by psychologists using statistical analysis, and computer scientists using feature engineering- and deep learning-based machine learning. We outlined present activities and additionally suggested new potential tasks based on recent research. Finally, we highlight several research limitations and suggest a number of future options, including the use of cutting-edge learning strategies, interpretable intention understanding, temporal detection, and proactive conversational intervention. Future suicidal thought detection will most likely take place mostly through online social content. In order to identify online messages containing suicidal ideation and to prevent suicide, it is crucial to create new techniques that can bridge the gap between clinical mental health detection and automatic machine detection.

Discussion and Future Work:

Suicidal ideation identification has been the subject of numerous preliminary studies, which have been greatly aided by manual feature engineering and DNN-based representation learning algorithms. However, there are several limits to the existing research, and there are still many obstacles to overcome.

A. Limitations:

1. Data Deficiency**:** Lack of data is the most important problem in the field of contemporary study. The majority of current approaches use supervised learning strategies, which necessitate manual annotation. To facilitate additional research, there aren't enough labeled data, though. For instance, there isn't any data with social links or multi-aspects, and labeled data with fine-grained suicide risk only contain a small number of examples.

2. Annotation Bias: To get the truth, there isn't much evidence to back up the suicide move. As a result, the most recent data were gathered through human labeling using some specified annotation rules. Annotations based on crowd-sourcing may produce biased labels. Shing et al.'s [13] request for expert labeling resulted in a small number of labeled occurrences. Regarding the demographic information, the quality of the data on suicides is alarming, and mortality estimation only includes general deaths rather than suicides11. Some incidents are incorrectly labeled as accidents or deaths with unclear motives.

3. Data Imbalance: Only a small portion of popular social media posts have suicidal intentions. However, rather than considering it as an unevenly distributed set of data, the majority of research constructed datasets in an even fashion to gather roughly balanced positive and negative samples.

4. Lack of Intention Understanding: Suicidal intent was not well understood by the existing statistical learning method. Complex psychological factors have a role in attempted suicide. Mainstream approaches, however, concentrate on choosing characteristics or employing sophisticated neural architectures to improve the performance of prediction. Machine learning methodologies discovered statistical hints from the phenomenology of suicidal posts in social material. However, despite including the psychology of suicide, they were unable to rationalize the risk factors.

B. Future Work:

1. Emerging Learning Techniques: The development of deep learning methodologies has accelerated studies on the identification of suicidal ideation. newer methods of instruction such as attention mechanism and graph neural networks, can be introduced for suicide text representation learning. You can also use different learning models including reinforcement learning, adversarial training, and transfer learning. For instance, generative adversarial networks can be utilized to produce adversarial samples for data augmentation, and knowledge of the mental health detection domain can be transferred for the detection of suicidal ideation.
Suicidal ideation-related postings are more likely to be found in the long tail of post categories on social networking sites. Few-shot learning can be used to train on a small sample of posts from the enormous social corpus that have been tagged as having suicidal ideation in order to achieve effective identification in the unbalanced distribution of real-world scenarios.

2. Understanding and Interpretability of Suicidal Intention: Suicide is associated with a variety of factors, including mental health, economic recessions, the number of guns, seasonal patterns of light, divorce laws, media coverage of suicide, and alcohol usage 12. Suicidal intention can be better understood, which can act as a guide for successful detection and intervention. Deep learning models can be given commonsense thinking by, for instance, adding external knowledge bases on suicide, according to a new research path.
An accurate prediction model can be learned using deep learning techniques. This would be a black-box model, though. New interpretable models should be created in order to better comprehend people's suicide intentions and to have a trustworthy prognosis.

3. Detection of Temporal Suicidal Thoughts: Another approach is to look for suicidal ideation in the data stream while taking the temporal information into account. Stress, sadness, suicidal thoughts, and suicidal plans are just a few of the steps that might lead to a

ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10140847

suicide attempt. The temporal trajectory of people's posts can be modeled in order to track changes in mental health and identify early indicators of suicide intent.

4. Proactive Intervention in Conversation: Intervention and prevention are the ultimate goals of the detection of suicidal ideation. To enable proactive intervention, very little work is done. When suicidal identification and crisis management are combined, Proactive Suicide Prevention Online (PSPO) [105] offers a fresh viewpoint. Conversations are a useful method. A possible technical option to enable prompt intervention for suicidal ideation is automatic response creation. Techniques for natural language production can be used to create counseling replies to ease depressed or suicidal individuals' thoughts. Conversational suicide intervention can also be used in reinforcement learning. Online volunteers and laypeople will intervene to remark on the initial posts and convince attempters to give up their suicidality after suicide attempters publish suicide messages (as the initial state). do nothing, respond to the remarks, or help someone who is suicidal. As a reward, the performance of a suicide attempt will be measured to determine a score. The conversational suicide intervention employs a policy gradient to help agents develop responses with the greatest potential for rewarding persons who are considering suicide.

VIII. References

[1] S. Hinduja and J.W.Patchin, "Bullying, cyberbullying, and suicide," Archives of suicide research, vol. 14, no. 3, pp. 206–221,2010.

[2] J.Joo, S.Hwang, and J.J.Gallo, "Death ideation and suicidal ideation in a community sample who do not meet criteria for major depression, "Crisis,2016.

[3] M.J.Vioule, B.Moulahi, J.Aze´, and S.Bringay, "Detection of suicide-related posts in twitter data streams," IBM Journal of Research and Development, vol.62, no.1, pp.7:1–7:12,2018..

[4] A.J.Ferrari, R.E.Norman, G.Freedman, A.J.Baxter, J.E.Pirkis,
M. G. Harris, A. Page, E. Carnahan, L. Degenhardt, T. Vos, et al., "The burden attributable to mental and substance use disorders as risk factors for suicide: findings from the global burden of disease study 2010," PloSone, vol.9, no.4, p.e91936,2014.

[5] R. C. O'Connor and M. K. Nock, "The psychology of suicidal behavior, "The Lancet Psychiatry, vol.1, no.1, pp.73–85,2014.

[6] J.Lopez-Castroman, B.Moulahi, J.Aze´, S.Bringay, J.Deninotti,
S.Guillaume, and E. Baca-Garcia, "Mining social networks to improve suicide prevention: As a coping review, "Journal of Neuroscience Research, 2019.

[7] C.M.McHugh, A.Corderoy, C.J.Ryan, I.B.Hickie, and
M. M. Large, "Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity, and positive predictive value, "BJ Psych open, vol.5,no.2,2019.

[8] G. Kassen, A. Kudaibergenova, A. Mukasheva, D. Yertargynkyzy, and K. Moldassan, "Behavioral risk factors for suicide among adolescent school children, " Elementary Education Online, vol. 19,no.1, pp.66–77,2019.

[9] V. Venek, S. Scherer, L.-P. Morency, J. Pestian, et al., "Adolescent suicidal risk assessment in clinician-patient interaction," IEEE Transactions on Affective Computing, vol.8, no.2, pp.204–215,2017.

[10]. D. Delgado-Gomez, H. Blasco-Fontecilla, A. A. Alegria, T. Legido-Gil, A.Artes-Rodriguez, and E.Baca-Garcia, "Improving the accuracy of suicide attempter classification, "Artificial intelligence in medicine, vol.52, no.3, pp.165–168,2011.

[11] G. Liu, C.Wang, K. Peng, H. Huang, Y. Li, and W. Cheng, "SocInf: Membership inference attacks on social media health data with machine learning, " IEEE Transactions on Computational Social Systems, 2019.

[12] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," Internet Interventions, vol.2, no.2, pp.183–188,2015.

[13] H.-C.Shing, S.Nair, A.Zirikly, M.Friedenberg, H.Daume´III, and P. Resnik, "Expert, crowd-sourced, and machine assessment of suicide risk via online postings," in Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, 2018, pp.25–36.

[14] F. Ren, X. Kang, and C. Quan, "Examining accumulated emotional traits in suicide blogs with an emotion topic model," IEEE Journal of Biomedical and Health Informatics, vol.20, no. 5, pp. 1384–1396,2016.