Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications

Article *in* IEEE Transactions on Computational Social Systems · September 2020

DOI: 10.1109/TCSS.2020.3021467

6 authors, including:

Shaoxiong Ji
University of Helsinki
60 PUBLICATIONS **3,970** CITATIONS
SEE PROFILE

Shirui Pan
Griffith University
455 PUBLICATIONS **28,245** CITATIONS
SEE PROFILE

Xue Li
Beijing Jiaotong University
317 PUBLICATIONS **7,939** CITATIONS
SEE PROFILE

Erik Cambria
Nanyang Technological University
643 PUBLICATIONS **54,176** CITATIONS
SEE PROFILE

Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications

Shaoxiong Ji, Shirui Pan, *Member, IEEE,* Xue Li,
Erik Cambria, *Senior Member, IEEE,* Guodong Long, and Zi Huang

***Abstract*—Suicide is a critical issue in modern society. Early detection and prevention of suicide attempts should be addressed to save people's life. Current suicidal ideation detection methods include clinical methods based on the interaction between social workers or experts and the targeted individuals and machine learning techniques with feature engineering or deep learning for automatic detection based on online social contents. This paper is the first survey that comprehensively introduces and discusses the methods from these categories. Domain-specific applications of suicidal ideation detection are reviewed according to their data sources, i.e., questionnaires, electronic health records, suicide notes, and online user content. Several specific tasks and datasets are introduced and summarized to facilitate further research. Finally, we summarize the limitations of current work and provide an outlook of further research directions.**

***Index Terms*—Suicidal ideation detection, social content, feature engineering, deep learning.**

I. Introduction

MENTAL health issues, such as anxiety and depression, are becoming increasingly concerned in modern society, as they turn out to be especially severe in developed countries and emerging markets. Severe mental disorders without effective treatment can turn to suicidal ideation or even suicide attempts. Some online posts contain much negative information and generate problematic phenomena such as cyberstalking and cyberbullying. Consequences can be severe and risky since such lousy information is often engaged in some form of social cruelty, leading to rumors or even mental damage. Research shows that there is a link between cyberbullying and suicide [1]. Victims overexposed to too many negative messages or events may become depressed and desperate; even worse, some may commit suicide.

The reasons that people commit suicide are complicated. People with depression are highly likely to commit suicide, but many without depression can also have suicidal thoughts [2]. According to the American Foundation for Suicide Prevention (AFSP), suicide factors fall under three categories: health factors, environmental factors, and historical factors [3]. Ferrari et al. [4] found that mental health issues and substance use disorders are attributed to the factors of suicide. O'Connor and Nock [5] conducted a thorough review of the psychology of suicide and summarized psychological risks as personality and individual differences, cognitive factors, social factors, and negative life events.

Suicidal Ideation Detection (SID) determines whether the person has suicidal ideation or thoughts by given tabular data of a person or textual content written by a person. Due to the advances in social media and online anonymity, an increasing number of individuals turn to interact with others on the Internet. Online communication channels are becoming a new way for people to express their feelings, suffering, and suicidal tendencies. Hence, online channels have naturally started to act as a surveillance tool for suicidal ideation, and mining social content can improve suicide prevention [6]. Strange social phenomena are emerging, e.g., online communities reaching an agreement on self-mutilation and copycat suicide. For example, a social network phenomenon called the "Blue Whale Game"[1] in 2016 uses many tasks (such as self-harming) and leads game members to commit suicide in the end. Suicide is a critical social issue and takes thousands of lives every year. Thus, it is necessary to detect suicidality and to prevent suicide before victims end their life. Early detection and treatment are regarded as the most effective ways to prevent potential suicide attempts.

Potential victims with suicidal ideation may express their thoughts of committing suicide in fleeting thoughts, suicide plans, and role-playing. Suicidal ideation detection is to find out these risks of intentions or behaviors before tragedy strikes. A meta-analysis conducted by McHugh et al. [7] shown statistical limitations of ideation as a screening tool, but also pointed out that people's expression of suicidal ideation represents their psychological distress. Effective detection of early signals of suicidal ideation can identify people with suicidal thoughts and open a communication portal to let social workers mitigate their mental issues. The reasons for suicide are complicated and attributed to a complex interaction of many factors [5], [8]. To detect suicidal ideation, many researchers conducted psychological and clinical studies [9] and classified responses of questionnaires [10]. Based on their social media data, artificial intelligence (AI) and machine learning techniques can predict people's likelihood of suicide [11], which can better understand people's intentions and pave the way for early intervention. Detection on social content focuses on feature engineering [12], [13], sentiment analysis [14], [15], and deep learning [16], [17], [18]. Those methods generally require heuristics to select features or design artificial neural network architectures for learning rich representation. The research trend focuses on selecting more useful features from people's health records and developing neural architectures to understand the language with suicidal ideation better. Mobile technologies have been studied and applied to suicide prevention, for example, the mobile suicide intervention application iBobbly [19] developed by the Black Dog Institute[2].

S. Ji is with Aalto University, Finland and The University of Queensland, Australia. E-mail: shaoxiong.ji@aalto.fi

X. Li, and Z. Huang are with The University of Queensland, Australia. E-mail: {xueli; huang}@itee.uq.edu.au

S. Pan is with Monash University, Australia. E-mail: shirui.pan@monash.edu

E. Cambria is with Nanyang Technological University, Singapore. E-mail: cambria@ntu.edu.sg

G. Long is with University of Technology Sydney, Australia. E-mail: guodong.long@uts.edu.au

[1] https://thesun.co.uk/news/worldnews/3003805

[2] https://blackdoginstitute.org.au/research/digital-dog/programs/ibobbly-app

Many other suicide prevention tools integrated with social networking services have also been developed, including Samaritans Radar[3] and Woebot[4]. The former was a Twitter plugin that was later discontinued because of privacy issues. For monitoring alarming posts. The latter is a Facebook chatbot based on cognitive behavioral therapy and natural language processing (NLP) techniques for relieving people's depression and anxiety. Applying cutting-edge AI technologies for suicidal ideation detection inevitably comes with privacy issues [20] and ethical concerns [21]. Linthicum et al. [22] put forward three ethical issues, including the influence of bias on machine learning algorithms, the prediction on time of suicide act, and ethical and legal questions raised by false positive and false negative prediction. It is not easy to answer ethical questions for AI as these require algorithms to reach a balance between competing values, issues, and interests [20]. AI has been applied to solve many challenging social problems. Detection of suicidal ideation with AI techniques is one of the potential applications for social good and should be addressed to improve people's wellbeing meaningfully. The research problems include feature selection on tabular and text data and representation learning on natural language. Many AI-based methods have been applied to classify suicide risks. However, there remain some challenges.

There is a limited number of benchmarks for training and evaluating suicidal ideation detection. AI-powered models sometimes learn statistical clues, but fail to understand people's intention. Moreover, many neural models are lack of interpretability. This survey reviews suicidal ideation detection methods from the perspective of AI and machine learning and specific domain applications with social impact. The categorization from these two perspectives is shown in Fig. 1. This paper provides a comprehensive review of the increasingly important field of suicidal ideation detection with machine learning methods. It proposes a summary of current research progress and an outlook of future work. The contributions of our survey are summarized as follows.

- To the best of our knowledge, this is the first survey that conducts a comprehensive review of suicidal ideation detection, its methods, and its applications from a machine learning perspective.
- We introduce and discuss the classical content analysis and modern machine learning techniques, plus their application to questionnaires, EHR data, suicide notes, and online social content.
- We enumerate existing and less explored tasks and discuss their limitations. We also summarize existing datasets and provide an outlook of future research directions in this field.

The remainder of the paper is organized as follows: methods and applications are introduced and summarized in Section II and Section III, respectively; Section IV enumerates specific tasks and some datasets; finally, we have a discussion and propose some future directions in Section V.

II. Methods and Categorization

Suicide detection has drawn the attention of many researchers due to an increasing suicide rate in recent years and has been studied extensively from many perspectives. The research techniques used to examine suicide also span many fields and methods, for example, clinical methods with patient-clinic interaction [9] and automatic detection from user-generated content (mainly text) [12], [17]. Machine learning techniques are widely applied for automatic detection.

Traditional suicide detection relies on clinical methods, including self-reports and face-to-face interviews. Venek et al. [9] designed a five-item ubiquitous questionnaire for the assessment of suicidal risks and applied a hierarchical classifier on the patients' response to determine their suicidal intentions. Through face-to-face interaction, verbal and acoustic information can be utilized. Scherer [23] investigated the prosodic speech characteristics and voice quality in a dyadic interview to identify suicidal and non-suicidal juveniles. Other clinical methods examine resting state heart rate from converted sensing signals [24], classify functional magnetic resonance imaging-based neural representations of death- and life-related words [25], and event-related instigators converted from EEG signals [26]. Another aspect of clinical treatment is the understanding of the psychology behind suicidal behavior [5], which, however, relies heavily on the clinician's knowledge and face-to-face interaction. Suicide risk assessment scales with clinical interview can reveal informative cues for predicting suicide [27]. Tan et al. [28] conducted an interview and survey study in Weibo, a Twitter-like service in China, to explore the engagement of suicide attempters with intervention by direct messages.

A. Content Analysis

Users' post on social websites reveals rich information and their language preferences. Through exploratory data analysis on the user-generated content can have an insight into language usage and linguistic clues of suicide attempters. The detailed analysis includes lexicon-based filtering, statistical linguistic features, and topic modeling within suicide-related posts.

Suicide-related keyword dictionary and lexicon are manually built to enable keyword filtering [29], [30] and phrases filtering [31]. Suicide-related keywords and phrases include "kill", "suicide", "feel alone", "depressed", and "cutting myself". Vioulès et al. [3] built a point-wise mutual information symptom lexicon using an annotated Twitter dataset. Gunn and Lester [32] analyzed posts from Twitter in the 24 hours before the death of a suicide attempter. Coppersmith et al. [33] analyzed the language usage of data from the same platform. Suicidal thoughts may involve strong negative feelings, anxiety, and hopelessness, or other social factors like family and friends. Ji et al. [17] performed word cloud visualization and topics modeling over suicide-related content and found that suicide-related discussion covers personal and social issues. Colombo et al. [34] analyzed the graphical characteristics of connectivity and communication in the Twitter social network. Coppersmith et al. [35] provided an exploratory analysis of language patterns and emotions on Twitter. Other methods and techniques include Google Trends analysis for suicide risk monitoring [36], the reply bias assessment through linguistic clues [37], human-machine hybrid method for analysis of the language effect of social support on suicidal ideation risk [38], social media content detection, and speech patterns analysis [39].

B. Feature Engineering

The goal of text-based suicide classification is to determine whether candidates, through their posts, have suicidal ideations. Machine learning methods and NLP have also been applied in this field.

[3] https://samaritans.org/about-samaritans/research-policy/internet-suicide/samaritans-radar

[4] https://woebot.io

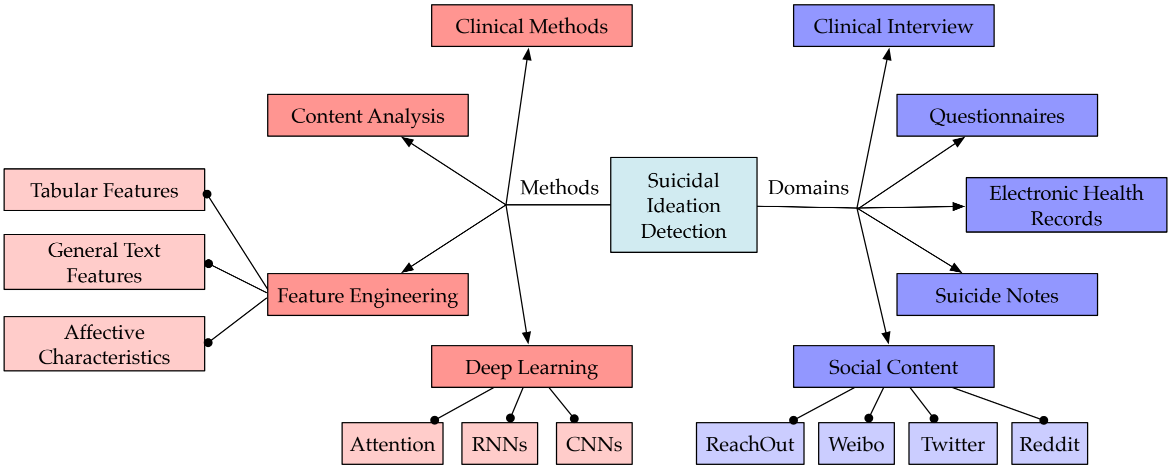
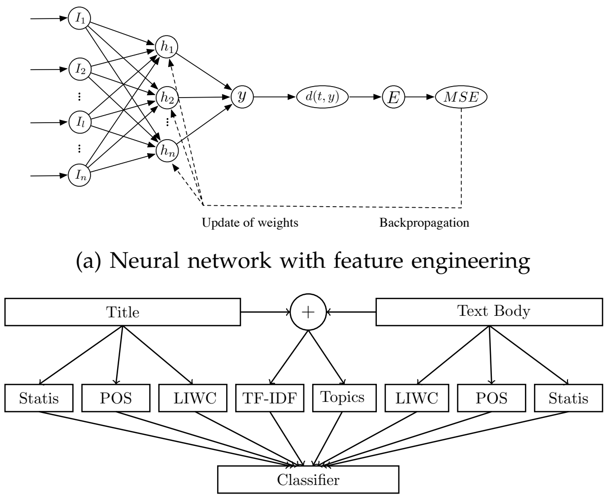


Fig. 1: The categorization of suicide ideation detection: methods and domains. The left part represents method categorization, while the right part shows the categories of domains. The arrow and solid point indicate subcategories.

1) Tabular Features: Tabular data for suicidal ideation detection consist of questionnaire responses and structured statistical information extracted from websites. Such structured data can be directly used as features for classification or regression. Masuda et al. [40] applied logistic regression to classify suicide and control groups based on users' characteristics and social behavior variables. The authors found variables such as community number, local clustering coefficient, and homophily have a more substantial influence on suicidal ideation in an SNS of Japan. Chattopadhyay [41] applied Pierce Suicidal Intent Scale (PSIS) to assess suicide factors and conducted regression analysis. Questionnaires act as a good source of tabular features. Delgado-Gomez et al. [42] used the international personality disorder examination screening questionnaire and the Holmes-Rahe social readjustment rating scale. Chattopadhyay [43] proposed to apply a multilayer feed-forward neural network, as shown in Fig. 2a to classify suicidal intention indicators according to Beck's suicide intent scale.

2) General Text Features: Another direction of feature engineering is to extract features from unstructured text. The main features consist of N-gram features, knowledge-based features, syntactic features, context features, and class-specific features [44]. Abboute et al. [45] built a set of keywords for vocabulary feature extraction within nine suicidal topics. Okhapkina et al. [46] built a dictionary of terms about suicidal content. They introduced term frequency-inverse document frequency (TF-IDF) matrices for messages and a singular value decomposition (SVD) for matrices. Mulholland and Quinn [47] extracted vocabulary and syntactic features to build a classifier to predict the likelihood of a lyricist's suicide. Huang et al. [48] built a psychological lexicon dictionary by extending HowNet (a commonsense word collection) and used a support vector machines (SVM) to detect cybersuicide in Chinese microblogs. The topic model [49] is incorporated with other machine learning techniques for identifying suicide in Sina Weibo. Ji et al. [17] extract several informative sets of features, including statistical, syntactic, linguistic inquiry and word count (LIWC), word embedding, and topic features, and then put the extracted features into classifiers as shown in Fig. 2b, where four traditional supervised classifiers are compared. Shing et al. [13] extracted several features as a bag of words (BoWs), empath, readability, syntactic features, topic model posteriors, word embeddings, linguistic inquiry and word count, emotion features and mental disease lexicon.

Models for suicidal ideation detection with feature engineering include SVM [44], artificial neural networks (ANN) [50] and conditional random field (CRF) [51]. Tai et al. [50] selected several features, including the history of suicide ideation and self-harm behavior, religious belief, family status, mental disorder history of candidates, and their family. Pestian et al. [52] compared the performance of different multivariate techniques with features of word counts, POS, concepts, and readability scores. Similarly, Ji et al. [17] compared four classification methods of logistic regression, random forest, gradient boosting decision tree, and XGBoost. Braithwaite et al. [53] validated machine learning algorithms can effectively identify high suicidal risk.



(a) Neural network with feature engineering

(b) Classifier with feature engineering

Fig. 2: Illustrations of methods with feature engineering

3) Affective Characteristics: Affective characteristics are among the most distinct differences between those who attempt suicide and healthy individuals, which has drawn considerable attention from both computer scientists and mental health researchers. To detect the emotions in suicide notes, Liakata et al. [51] used manual emotion categories, including anger, sorrow, hopefulness, happiness/peacefulness, fear, pride, abuse, and forgiveness. Wang et al. [44] employed combined characteristics of both factual (2 categories) and sentimental aspects (13 categories) to discover fine-grained sentiment analysis. Similarly, Pestian et al. [52] identified emotions of abuse, anger, blame, fear, guilt, hopelessness, sorrow, forgiveness, happiness, peacefulness, hopefulness, love, pride, thankfulness, instructions, and information. Ren et al. [14] proposed a complex emotion topic model and applied it to analyze accumulated emotional traits in suicide blogs and to detect suicidal intentions from a blog stream. Specifically, the authors studied accumulate emotional traits, including emotion accumulation, emotion covariance, and emotion transition among eight basic emotions of joy, love, expectation, surprise, anxiety, sorrow, anger, and hate with a five-level intensity.

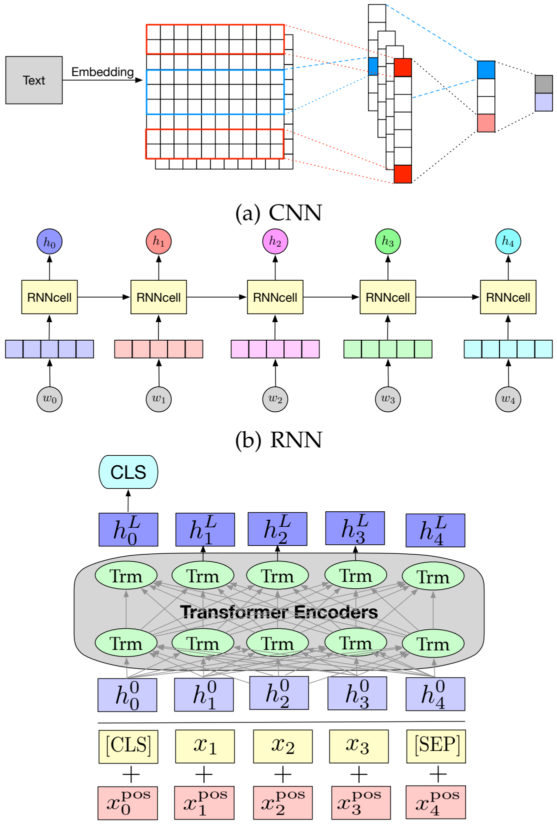
C. Deep Learning

Deep learning has been a great success in many applications, including computer vision, NLP, and medical diagnosis. In the field of suicide research, it is also an important method for automatic suicidal ideation detection and suicide prevention. It can effectively learn text features automatically without sophisticated feature engineering techniques. At the same time, some also take extracted features into deep neural networks; for example, Nobles et al. [54] fed psycholinguistic features and word occurrence into the multilayer perceptron (MLP). Popular deep neural networks (DNNs) include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT), as shown in Fig. 3a, 3b and 3c. Natural language text is usually embedded into distributed vector space with popular word embedding techniques such as word2vec [55] and GloVe [56]. Shing et al. [13] applied user-level CNN with the filter size of 3, 4, and 5 to encode users' posts. Long short-term memory (LSTM) network, a popular variant of RNN, is applied to encode textual sequences and then process for classification with fully connected layers [17].

Recent methods introduce other advanced learning paradigms to integrate with DNNs for suicidal ideation detection. Ji et al. [57] proposed model aggregation methods for updating neural networks, i.e., CNNs and LSTMs, targeting to detect suicidal ideation in private chatting rooms. However, decentralized training relies on coordinators in chatting rooms to label user posts for supervised training, which can only be applied to minimal scenarios. One possible better way is to use unsupervised or semi-supervised learning methods. Benton et al. [16] predicted suicide attempt and mental health with neural models under the framework of multi-task learning by predicting the gender of users as an auxiliary task. Gaur et al. [58] incorporated external knowledge bases and suicide-related ontology into a text representation and gained an improved performance with a CNN model. Coppersmith et al. [59] developed a deep learning model with GloVe for word embedding, bidirectional LSTM for sequence encoding, and self-attention mechanism for capturing the most informative subsequence. Sawhney et al. [60] used LSTM, CNN, and RNN for suicidal ideation detection. Similarly, Tadesse et al. [61] employed LSTM-CNN model. Ji et al. [62] proposed an attentive relation network with LSTM and topic modeling for encoding text and risk indicators.

In the 2019 CLPsych Shared Task [63], many popular DNN architectures were applied. Hevia et al. [64] evaluated the effect of pretraining using different models, including GRU-based RNN. Morales et al. [65] studied several popular deep learning models such as CNN, LSTM, and Neural Network Synthesis (NeuNetS). Matero et al. [66] proposed dual-context model using hierarchically attentive RNN, and BERT.

Another sub-direction is the so-called hybrid method, which cooperates minor feature engineering with representation learning techniques. Chen et al. [67] proposed a hybrid classification model of the behavioral model and the suicide language model. Zhao et al. [68] proposed the D-CNN model taking word embedding and external tabular features as inputs for classifying suicide attempters with depression.



(a) CNN

(b) RNN

(c) BERT

Fig. 3: Deep neural networks for suicidal ideation detection

D. Summary

The popularization of machine learning has facilitated research on suicidal ideation detection from multi-modal data and provided a promising way for effective early warning. Current research focuses on text-based methods by extracting features and deep learning for automatic feature learning. Researchers widely use many canonical NLP features such as TF-IDF, topics, syntactic, affective characteristics, and readability, and deep learning models like CNN and LSTM. Those methods, especially deep neural networks with automatic feature learning, boosted predictive performance and preliminary success on suicidal intention understanding. However, some methods may only learn statistical cues and lack of commonsense. The recent work [58] incorporated external knowledge using knowledge bases and suicide ontology for knowledge-aware suicide risk assessment. It took a remarkable step towards knowledge-aware detection.

TABLE I: Categorization of methods for suicidal ideation detection

Category	Publications	Methods	Inputs
Feature Engineering	Ji et al. [17]	Word counts, POS, LIWC, TF-IDF + classifiers	Text
	Masuda et al. [40]	Multivariate/univariate logistic regression	Characteristics variables
	Delgado-Gomez et al. [42]	International personal disorder examination screening questionnaire	Questionnaire responses
	Mulholland et al. [47]	Vocabulary features, syntactic features, semantic class features, N-gram	lyrics
	Okhapkina et al. [46]	Dictionary, TF-IDF + SVD	Text
	Huang et al. [48]	Lexicon, syntactic features, POS, tense	Text
	Pestian et al. [52]	Word counts, POS, concepts and readability score	Text
	Tai et al. [50]	self-measurement scale + ANN	Self-measurement forms
	Shing et al. [13]	BoWs, empath, readability, syntactic, topic, LIWC, emotion, lexicon	Text
Deep Learning	Zhao et al. [68]	Word embedding, tabular features, D-CNN	Text+external information
	Shing et al. [13]	Word embedding, CNN, max pooling	Text
	Ji et al. [17]	Word embedding, LSTM, max pooling	Text
	Bento et al. [16]	Multi-task learning, neural networks	Text
	Nobles et al. citenobles2018identification	MLP, psycholinguistic features, word occurrence	Text
	Hevia et al. [64]	Pretrained GRU, word embedding, document embedding	Text
	Morales et al. [65]	CNN, LSTM, NeuNetS, word embedding	Text
	Matero et al. [66]	Dual-context, BERT, GRU, attention, user-factor adaptation	Text
	Gaur et al. [58]	CNN, knowledge base, ConceptNet embedding	Text
	Coppersmith et al. [59]	GloVe, BiLSTM, self attention	Text
	Ji et al. [62]	Relation network, LSTM, attention, lexicon	Text
	Tadesse et al. [61]	LSTM, CNN, word embedding	Text

III. Applications on Domains

Many machine learning techniques have been introduced for suicidal ideation detection. The relevant extant research can also be viewed according to the data source. Specific applications cover a wide range of domains, including questionnaires, electronic health records (EHRs), suicide notes, and online user content. Fig. 4 shows some examples of data source for suicidal ideation detection, where Fig. 4a lists selected questions of the "International Personal Disorder Examination Screening Questionnaire" (IPDE-SQ) adapted from [42], Fig. 4b are selected patient's records from [69], Fig. 4c is a suicide note from a website[5], and Fig. 4d is a tweet and its corresponding comments from Twitter.com. Nobles et al. [54] identified suicide risk using text messages. Some researchers also developed softwares for suicide prevention. Berrouiguet et al. [70] developed a mobile application for health status self report. Meyer et al. [71] developed an e-PASS Suicidal Ideation Detector (eSID) tool for medical practitioners. Shah et al. [72] utilized social media videos and studied multimodal behavioral markers.

A. Questionnaires

Mental disorder scale criteria such as DSM-IV[6] and ICD-10[7], and the IPDE-SQ provides good tool for evaluating an individual's mental status and their potential for suicide. Those criteria and examination metrics can be used to design questionnaires for self-measurement or face-to-face clinician-patient interview. To study the assessment of suicidal behavior, Delgado-Gomez et al. [10] applied and compared the IPDE-SQ and the "Barrat's Impulsiveness Scale"(version 11, BIS-11) to identify people likely to attempt suicide. The authors also conducted a study on individual items from those two scales. The BIS-11 scale has 30 items with 4-point ratings, while the IPDE-SQ in DSM-IV has 77 true-false screening questions. Further, Delgado-Gomez et al. [42] introduced the "Holmes-Rahe Social Readjustment Rating Scale" (SRRS) and the IPDE-SQ as well to two comparison groups of suicide attempters and non-suicide attempters. The SRRS consists of 43 ranked life events of different levels of severity. Harris et al. [73] surveyed understanding suicidal individuals' online behaviors to assist suicide prevention. Sueki [74] conducted an online panel survey among Internet users to study the association between suicide-related Twitter use and suicidal behavior. Based on the questionnaire results, they applied several supervised learning methods, including linear regression, stepwise linear regression, decision trees, Lars-en, and SVMs, to classify suicidal behaviors.

B. Electronic Health Records

The increasing volume of electronic health records (EHRs) has paved the way for machine learning techniques for suicide attempter prediction. Patient records include demographical information and diagnosis-related history like admissions and emergency visits. However, due to the data characteristics such as sparsity, variable length of clinical series, and heterogeneity of patient records, many challenges remain in modeling medical data for suicide attempt prediction. Besides, the recording procedures may change because of the change of healthcare policies and the update of diagnosis codes.

There are several works of predicting suicide risk based on EHRs [75], [76]. Tran et al. [69] proposed an integrated suicide risk prediction framework with a feature extraction scheme, risk classifiers, and risk calibration procedure. Explicitly, each patient's clinical history is represented as a temporal image. Iliou et al. [77] proposed a data preprocessing method to boost machine learning techniques for suicide tendency prediction of patients suffering from mental disorders. Nguyen et al. [78] explored real-world administrative data of mental health patients from the hospital for short and medium-term suicide risk assessments. By introducing random forests, gradient boosting machines, and DNNs, the authors managed to deal with high dimensionality and redundancy issues of data. Although the previous method gained preliminary success, Iliou et al. [77] and Nguyen et al. [78] have a limitation on the source of data which focuses on patients with mental disorders in their historical records. Bhat and Goldman-Mellor [79] used an anonymized general EHR dataset to relax the restriction on patient's diagnosis-related history, and applied neural networks as a classification model to predict suicide attempters.

[5] https://paranorms.com/suicide-notes

[6] https://psychiatry.org/psychiatrists/practice/dsm

[7] https://apps.who.int/classifications/icd10/browse/2016/en

C. Suicide Notes

Suicide notes are the written notes left by people before committing suicide. They are usually written on letters and online blogs, and recorded in audio or video. Suicide notes provide material for NLP research. Previous approaches have examined suicide notes using content analysis [52], sentiment analysis [80], [44], and emotion detection [51]. Pestian et al. [52] used transcribed suicide notes with two groups of completers and elicitors from people who have a personality disorder or potential morbid thoughts. White and Mazlack [81] analyzed word frequencies in suicide notes using a fuzzy cognitive map to discern causality. Liakata et al. [51] employed machine learning classifiers to 600 suicide messages with varied length, different readability quality, and multi-class annotations.

Emotion in text provides sentimental cues of suicidal ideation understanding. Desmet et al. [82] conducted a fine-grained emotion detection on suicide notes of 2011 i2b2 task. Wicentowski and Sydes [83] used an ensemble of maximum entropy classification. Wang et al. [44] and Kovačević et al. [84] proposed hybrid machine learning and rule-based method for the i2b2 sentiment classification task in suicide notes.

In the age of cyberspace, more suicide notes are now written in the form of web blogs and can be identified as carrying the potential risk of suicide. Huang et al. [29] monitored online blogs from MySpace.com to identify at-risk bloggers. Schoene and Dethlefs [85] extracted linguistic and sentiment features to identity genuine suicide notes and comparison corpus.

D. Online User Content

The widespread use of mobile Internet and social networking services facilitates people's expressing their life events and feelings freely. As social websites provide an anonymous space for online discussion, an increasing number of people suffering from mental disorders turn to seek for help. There is a concerning tendency that potential suicide victims post their suicidal thoughts on social websites like Facebook, Twitter, Reddit, and MySpace. Social media platforms are becoming a promising tunnel for monitoring suicidal thoughts and preventing suicide attempts [86]. Massive user-generated data provide a good source to study online users' language patterns. Using data mining techniques on social networks and applying machine learning techniques provide an avenue to understand the intent within online posts, provide early warnings, and even relieve a person's suicidal intentions.

Twitter provides a good source for research on suicidality. O'Dea et al. [12] collected tweets using the public API and developed automatic suicide detection by applying logistic regression and SVM on TF-IDF features. Wang et al. [87] further improved the performance with effective feature engineering. Shepherd et al. [88] conducted psychology-based data analysis for contents that suggests suicidal tendencies in Twitter social networks. The authors used the data from an online conversation called #dearmentalhealthprofessionals.

Another famous platform Reddit is an online forum with topic-specific discussions has also attracted much research interest for studying mental health issues [89] and suicidal ideation [37]. A community on Reddit called SuicideWatch is intensively used for studying suicidal intention [90], [17]. De Choudhury et al. [90] applied a statistical methodology to discover the transition from mental health issues to suicidality. Kumar et al. [91] examined the posting activity following the celebrity suicides, studied the effect of celebrity suicides on suicide-related contents, and proposed a method to prevent the high-profile suicides. Many pieces of research [48], [49] work on detecting suicidal ideation in Chinese microblogs. Guan et al. [92] studied user profile and linguistic features for estimating suicide probability in Chinese microblogs. There also remains some work using other platforms for suicidal ideation detection. For example, Cash et al. [93] conducted a study on adolescents' comments and content analysis on MySpace. Steaming data provides a good source for user pattern analysis. Vioulès et al. [3] conducted user-centric and post-centric behavior analysis and applied a martingale framework to detect sudden emotional changes in the Twitter data stream for monitoring suicide warning signs. Ren et al. [14] use the blog stream collected from public blog articles written by suicide victims to study the accumulated emotional information.

E. Summary

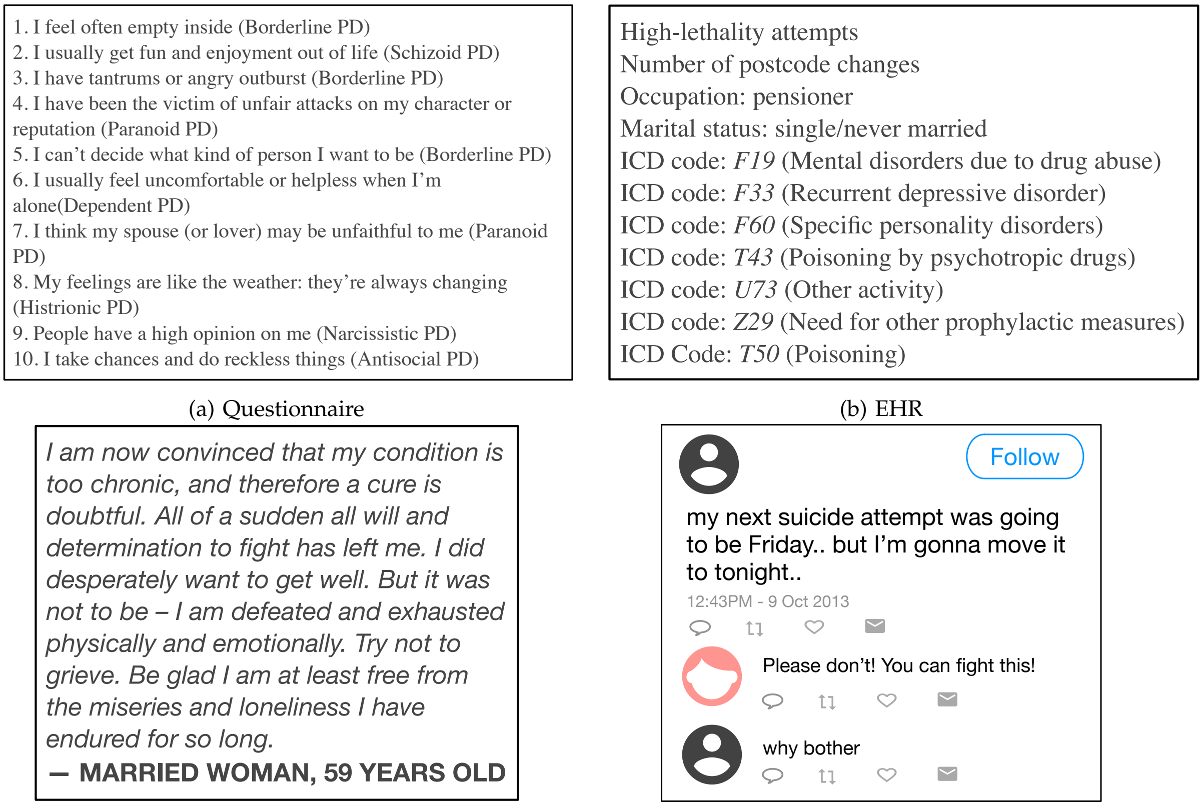
Applications of suicidal ideation detection mainly consist of four domains, i.e., questionnaires, electronic health records, suicide notes, and online user content. Table II gives a summary of categories, data sources, and methods. Among these four main domains, questionnaires and EHRs require self-report measurement or patient-clinician interactions and rely highly on social workers or mental health professions. Suicide notes have a limitation on immediate prevention, as many suicide attempters commit suicide in a short time after they write suicide notes. However, they provide a good source for content analysis and the study of suicide factors. The last online user content domain is one of the most promising ways of early warning and suicide prevention when empowered with machine learning techniques. With the rapid development of digital technology, user-generated content will play a more important role in suicidal ideation detection. Other forms of data, such as health data generated by wearable devices, can be very likely to help with suicide risk monitoring in the future.

TABLE II: Summary of studies on suicidal ideation detection from the views of intervention categories, data and methods

Categories	self-report examination [9] face-to-face suicide prevention [23] automatic SID II-B II-C III-C III-B III-D
Data	questionnaires III-A suicide notes III-C suicide blogs III-C electronic health records III-B online social texts III-D
Methods	clinical methods [24], [25], [26] mobile applications [19] content analysis II-A feature engineering II-B deep learning II-C
Critical issues	suicide factors [1], [2], [3], [4], [5] ethics [20], [21], [22] privacy [20]

IV. Tasks and Datasets

In this section, we summarize specific tasks in suicidal ideation detection and other suicide-related tasks about mental disorders. Some tasks, such as reasoning suicidal messages, generating a response, and suicide attempters detection on a social graph, may lack benchmarks for evaluation. However,



1. I feel often empty inside (Borderline PD)
2. I usually get fun and enjoyment out of life (Schizoid PD)
3. I have tantrums or angry outburst (Borderline PD)
4. I have been the victim of unfair attacks on my character or reputation (Paranoid PD)
5. I can't decide what kind of person I want to be (Borderline PD)
6. I usually feel uncomfortable or helpless when I'm alone(Dependent PD)
7. I think my spouse (or lover) may be unfaithful to me (Paranoid PD)
8. My feelings are like the weather: they're always changing (Histrionic PD)
9. People have a high opinion on me (Narcissistic PD)
10. I take chances and do reckless things (Antisocial PD)

(a) Questionnaire

High-lethality attempts
Number of postcode changes
Occupation: pensioner
Marital status: single/never married
ICD code: *F19* (Mental disorders due to drug abuse)
ICD code: *F33* (Recurrent depressive disorder)
ICD code: *F60* (Specific personality disorders)
ICD code: *T43* (Poisoning by psychotropic drugs)
ICD code: *U73* (Other activity)
ICD code: *Z29* (Need for other prophylactic measures)
ICD Code: *T50* (Poisoning)

(b) EHR

I am now convinced that my condition is too chronic, and therefore a cure is doubtful. All of a sudden all will and determination to fight has left me. I did desperately want to get well. But it was not to be – I am defeated and exhausted physically and emotionally. Try not to grieve. Be glad I am at least free from the miseries and loneliness I have endured for so long.

— MARRIED WOMAN, 59 YEARS OLD

(c) Suicide notes

my next suicide attempt was going to be Friday.. but I'm gonna move it to tonight..

12:43PM - 9 Oct 2013

Please don't! You can fight this!

why bother

(d) Tweets

Fig. 4: Examples of content for suicidal ideation detection

they are critical for effective detection. We propose these tasks together with the current research direction and call for contribution to these tasks from the research community. Meanwhile, an elaborate list of datasets for currently available tasks is provided, and some potential data sources are also described to promote the research efforts.

A. Tasks

1) Suicide Text Classification: The first task - suicide text classification can be viewed as a domain-specific application of general text classification, which includes binary and multi-class classification. Binary suicidality classification simply determines text with suicidal ideation or not, while multi-class suicidality classification conducts fine-grained suicide risk assessment. For example, some research divides suicide risk into five levels: no, low, moderate, and severe. Alternatively, it can also consider four types of class labels according to mental and behavioral procedures, i.e., non-suicidal, suicidal thoughts/wishes, suicidal intentions, and suicidal act/plan.

Another subtask is risk assessment by learning from multi-aspect suicidal posts. Adopting the definition of characteristics of suicidal messages, Gilat et al. [94] manually tagged suicidal posts with multi-aspect labels, including mental pain, cognitive attribution, and level of suicidal risk. Mental pain includes loss of control, acute loneliness, emptiness, narcissistic wounds, irreversibility loss of energy, and emotional flooding, scaled into $[0, 7]$. Cognitive attribution is the frustration of needs associated with interpersonal relationships, or there is no indication of attribution.

2) Reasoning Suicidal Messages: Massive data mining and machine learning algorithms have achieved remarkable outcomes by using DNNs. However, simple feature sets and classification models are not predictive enough to detect complicated suicidal intentions. Machine learning techniques require reasoning suicidal messages to have a more in-depth insight into suicidal factors and the innermost being from textual posts. This task aims to employ interpretable methods to investigate suicidal factors and incorporate them with commonsense reasoning, which may improve the prediction of suicidal factors. Specific tasks include automatic summarization of suicide factor, find an explanation of suicidal risk in mental pain and cognitive attribution aspects associated with suicide.

3) Suicide Attempter Detection: The two tasks mentioned above focus on a single text itself. However, the primary purpose of suicidal ideation detection is the identify suicide attempters. Thus, it is vital to achieving user-level detection, which consists of two folds, i.e., user-level multi-instance suicidality detection and suicide attempt detection on a graph. The former takes a bag of posts from individuals as input and conducts multi-instance learning over a bag of messages. The later identifies suicide attempters in a specific social graph built by the interaction between users in social networks. It considers the relationship between social users and can be regarded as a node classification problem in a graph.

4) Generating Response: The ultimate goal of suicidal ideation detection is intervention and suicide prevention. Many people with suicidal intentions tend to post their suffering at midnight. Another task is generating a thoughtful response for counseling potential suicidal victims to enable immediate social care and relieve their suicidal intention. Gilat et al., [94] introduced eight types of response strategies; they are emotional support, offering group support, empowerment, interpretation, cognitive change inducement, persuasion, advising, and referring. This task requires machine learning techniques, especially sequence-to-sequence learning, to have the ability to adopt effective response strategies to generate better response and eliminate people's suicidality. When social workers or volunteers go back online, this response generation technique can also generate hints for them to compose a thoughtful response.

5) Mental disorders and Self-harm risk: Suicidal ideation has a strong relationship with a mental health issue and self-harm risks. Thus, detecting severe mental disorders or self-harm risks is also an important task. Such works include depression detection [95], self-harm detection [96], stressful periods and stressor events detection [97], building knowledge graph for depression [98], and correlation analysis on depression and anxiety [99]. Corresponding subtasks in this field are similar to suicide text classification in Section IV-A1.

B. Datasets

1) Text Data:

a) Reddit: Reddit is a registered online community that aggregates social news and online discussions. It consists of many topic categories, and each area of interest within a topic is called a subreddit. A subreddit called "Suicide Watch"(SW)[8] is intensively used for further annotation as positive samples. Posts without suicidal content are sourced from other popular subreddits. Ji et al. [17] released a dataset with 3,549 posts with suicidal ideation. Shing et al. [13] published their UMD Reddit Suicidality Dataset with 11,129 users and total 1,556,194 posts and sampled 934 users for further annotation. Aladağ et al. [100] collected 508,398 posts using Google Cloud BigQuery, and manually annotated 785 posts.

b) Twitter: Twitter is a popular social networking service, where many users also talk about their suicidal ideation. Twitter is quite different from Reddit in post length, anonymity, and the way communication and interaction. Twitter user data with suicidal ideation and depression are collected by Coppersmith et al. [33]. Ji et al. [17] collected an imbalanced dataset of 594 tweets with suicidal ideation out of a total of 10,288 tweets. Vioulès et al. collected 5,446 tweets using Twitter streaming API [3], of which 2,381 and 3,065 tweets from the distressed users and normal users, respectively. However, most Twitter-based datasets are no longer available as per the policy of Twitter.

c) ReachOut: ReachOut Forum[9] is a peer support platform provided by an Australian mental health care organization. The ReachOut dataset [101] was firstly released in the CLPsych17 shared task. 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.

2) EHR: EHR data contain demographical information, admissions, diagnostic reports, and physician notes. A collection of electronic health records is from the California emergency department encounter and hospital admission. It contains 522,056 anonymous EHR records from California-resident adolescents. However, it is not public for access. Bhat and Goldman-Mellor [79] firstly used these records from 2006-2009 to predict the suicide attempt in 2010. Haerian et al. [102] selected 280 cases for evaluation from the Clinical Data Warehouse (CDW) and WebCIS database at NewYork Presbyterian Hospital/ Columbia University Medical Center. Tran et al. [69] studied emergency attendances with a least one risk assessment from the Barwon Health data warehouse. The selected dataset contains 7,746 patients and 17,771 assessments.

3) Mental Disorders: Mental health issues such as depression without effective treatment can turn into suicidal ideation. For the convenience of research on mental disorders, we also list several resources for monitoring mental disorders. The eRisk dataset of Early Detection of Signs of Depression [103] is released by the first task of the 2018 workshop at the Conference and Labs of the Evaluation Forum (CLEF), which focuses on early risk prediction on the Internet[10]. This dataset contains sequential text from social media. Another dataset is the Reddit Self-reported Depression Diagnosis (RSDD) dataset [95], which contains 9,000 diagnosed users with depression and approximately 107,000 matched control users.

V. Discussion and Future Work

Many preliminary works have been conducted for suicidal ideation detection, especially boosted by manual feature engineering and DNN-based representation learning techniques. However, current research has several limitations, and there are still great challenges for future work.

A. Limitations

a) Data Deficiency: The most critical issue of current research is data deficiency. Current methods mainly apply supervised learning techniques, which require manual annotation. However, there are not enough annotated data to support further research. For example, labeled data with fine-grained suicide risk only have limited instances, and there are no multi-aspect data and data with social relationships.

b) Annotation Bias: There is little evidence to confirm the suicide action to obtain ground truth. Thus, current data are obtained by manual labeling with some predefined annotation rules. The crowdsourcing-based annotation may lead to bias of labels. Shing et al. [13] asked experts for labeling but only obtained a limited number of labeled instances. As for the demographical data, the quality of suicide data is concerning, and mortality estimation is general death but not suicide[11]. Some cases are misclassified as accidents or death of undetermined intent.

c) Data Imbalance: Posts with suicidal intention account for a tiny proportion of massive social posts. However, most works built datasets in an approximately even manner to collect relatively balanced positive and negative samples rather than treating it as an ill-balanced data distributed.

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

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

[10] https://early.irlab.org

[11] World Health Organization, Preventing suicide: a global imperative, 2014. https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf

TABLE III: A summary of public datasets

Type	Publication	Source	Instances	Public Access
Text	Shing et al. [13]	Reddit	866/11,129	https://tinyurl.com/umd-suicidality
Text	Aladağ et al. [100]	Reddit	508,398	Request to the authors
Text	Coppersmith et al. [33]	Twitter	> 3,200	N.A
Text	Coppersmith et al. [104]	Twitter	1,711	https://tinyurl.com/clpsych-2015
Text	Vioulès et al. [3]	Twitter	5,446	N.A
Text	Milne et al. [101]	ReachOut	65,756	N.A
EHR	Bhat and Goldman-Mellor [79]	Hospital	522,056	N.A
EHR	Tran et al. [69]	Barwon Health	7,746	N.A
EHR	Haerian et al. [102]	CDW & WebCIS	280	N.A
Text	Pestian et al. [80]	Notes	1,319	2011 i2b2 NLP challenge
Text	Gaur et al. [58]	Reddit	500/2,181	Request to the authors
Text	eRisk 2018 [103]	Social networks	892	https://tec.citius.usc.es/ir/code/eRisk.html
Text	RSDD [95]	Reddit	9,000	http://ir.cs.georgetown.edu/resources/rsdd.html

d) Lack of Intention Understanding: The current statistical learning method failed to have a good understanding of suicidal intention. The psychology behind suicidal attempts is complex. However, mainstream methods focus on selecting features or using complex neural architectures to boost the predictive performance. From the phenomenology of suicidal posts in social content, machine learning methods learned statistical clues. However, they failed to reason over the risk factors by incorporating the psychology of suicide.

B. Future Work

1) Emerging Learning Techniques: The advances of deep learning techniques have boosted research on suicidal ideation detection. More emerging learning techniques, such as attention mechanism and graph neural networks, can be introduced for suicide text representation learning. Other learning paradigms, such as transfer learning, adversarial training, and reinforcement learning, can also be utilized. For example, knowledge of the mental health detection domain can be transferred for suicidal ideation detection, and generative adversarial networks can be used to generated adversarial samples for data augmentation.

In social networking services, posts with suicidal ideation are in the long tail of the distribution of different post categories. To achieve effective detection in the ill-balanced distribution of real-world scenarios, few-shot learning can be utilized to train on a few labeled posts with suicidal ideation among the large social corpus.

2) Suicidal Intention Understanding and Interpretability: Many factors are correlated to suicide, such as mental health, economic recessions, gun prevalence, daylight patterns, divorce laws, media coverage of suicide, and alcohol use[12]. A better understanding of suicidal intention can provide a guideline for effective detection and intervention. A new research direction is to equip deep learning models with commonsense reasoning, for example, by incorporating external suicide-related knowledge bases.

Deep learning techniques can learn an accurate prediction model. However, this would be a black-box model. In order to better understand people's suicidal intentions and have a reliable prediction, new interpretable models should be developed.

3) Temporal Suicidal Ideation Detection: Another direction is to detect suicidal ideation over the data stream and consider the temporal information. There exist several stages of suicide attempts, including stress, depression, suicidal thoughts, and suicidal plan. Modeling people's posts' temporal trajectory can effectively monitor the change of mental status and is essential for detecting early signs of suicidal ideation.

4) Proactive Conversational Intervention: The ultimate aim of suicidal ideation detection is intervention and prevention. Very little work is undertaken to enable proactive intervention. Proactive Suicide Prevention Online (PSPO) [105] provides a new perspective with the combination of suicidal identification and crisis management. An effective way is through conversations. Automatic response generation becomes a promising technical solution to enable timely intervention for suicidal thoughts. Natural language generation techniques can be utilized to generate counseling responses to comfort people's depression or suicidal ideation. Reinforcement learning can also be applied for conversational suicide intervention. After suicide attempters post suicide messages (as the initial state), online volunteers and lay individuals will take action to comment on the original posts and persuade attempters to give up their suicidality. The attempter may do nothing, reply to the comments, or get their suicidality relieved. A score will be defined by observing the reaction from a suicide attempter as a reward. The conversational suicide intervention uses a policy gradient for agents to generated responses with maximum rewards to best relieve people's suicidal thoughts.

VI. Conclusion

Suicide prevention remains an essential task in our modern society. Early detection of suicidal ideation is an important and effective way to prevent suicide. This survey investigates existing methods for suicidal ideation detection from a broad perspective which covers clinical methods like patient-clinician interaction and medical signal sensing; textual content analysis such as lexicon-based filtering and word cloud visualization; feature engineering including tabular, textual, and affective features; and deep learning-based representation learning like CNN- and LSTM-based text encoders. Four main domain-specific applications on questionnaires, EHRs, suicide notes, and online user content are introduced.

Psychological experts have conducted most work in this field with statistical analysis, and computer scientists with

[12] Report by Lindsay Lee, Max Roser, and Esteban Ortiz-Ospina in OurWorldInData.org, retrieved from https://ourworldindata.org/suicide

feature engineering based machine learning and deep learning-based representation learning. Based on current research, we summarized existing tasks and further proposed new possible tasks. Last but not least, we discuss some limitations of current research and propose a series of future directions, including utilizing emerging learning techniques, interpretable intention understanding, temporal detection, and proactive conversational intervention.

Online social content is very likely to be the main channel for suicidal ideation detection in the future. Therefore, it is essential to develop new methods, which can heal the schism between clinical mental health detection and automatic machine detection, to detect online texts containing suicidal ideation in the hope that suicide can be prevented.

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. Vioulès, B. Moulahi, J. Azé, 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," *PloS one*, vol. 9, no. 4, p. e91936, 2014.

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

[6] J. Lopez-Castroman, B. Moulahi, J. Azé, S. Bringay, J. Deninotti, S. Guillaume, and E. Baca-Garcia, "Mining social networks to improve suicide prevention: A scoping 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," *BJPsych 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 schoolchildren," *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. Daumé III, and P. Resnik, "Expert, crowdsourced, 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.

[15] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, pp. 1–47, 2018.

[16] A. Benton, M. Mitchell, and D. Hovy, "Multi-task learning for mental health using social media text," in *EACL*. Association for Computational Linguistics, 2017.

[17] S. Ji, C. P. Yu, S.-f. Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," *Complexity*, 2018.

[18] S. Ji, G. Long, S. Pan, T. Zhu, J. Jiang, and S. Wang, "Detecting suicidal ideation with data protection in online communities," in *24th International Conference on Database Systems for Advanced Applications (DASFAA)*. Springer, Cham, 2019, pp. 225–229.

[19] J. Tighe, F. Shand, R. Ridani, A. Mackinnon, N. De La Mata, and H. Christensen, "Ibobbly mobile health intervention for suicide prevention in australian indigenous youth: a pilot randomised controlled trial," *BMJ open*, vol. 7, no. 1, p. e013518, 2017.

[20] N. N. G. de Andrade, D. Pawson, D. Muriello, L. Donahue, and J. Guadagno, "Ethics and artificial intelligence: suicide prevention on facebook," *Philosophy & Technology*, vol. 31, no. 4, pp. 669–684, 2018.

[21] L. C. McKernan, E. W. Clayton, and C. G. Walsh, "Protecting life while preserving liberty: Ethical recommendations for suicide prevention with artificial intelligence," *Frontiers in psychiatry*, vol. 9, p. 650, 2018.

[22] K. P. Linthicum, K. M. Schafer, and J. D. Ribeiro, "Machine learning in suicide science: Applications and ethics," *Behavioral sciences & the law*, vol. 37, no. 3, pp. 214–222, 2019.

[23] S. Scherer, J. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 709–713.

[24] D. Sikander, M. Arvaneh, F. Amico, G. Healy, T. Ward, D. Kearney, E. Mohedano, J. Fagan, J. Yek, A. F. Smeaton *et al.*, "Predicting risk of suicide using resting state heart rate," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2016, pp. 1–4.

[25] M. A. Just, L. Pan, V. L. Cherkassky, D. L. McMakin, C. Cha, M. K. Nock, and D. Brent, "Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth," *Nature human behaviour*, vol. 1, no. 12, p. 911, 2017.

[26] N. Jiang, Y. Wang, L. Sun, Y. Song, and H. Sun, "An erp study of implicit emotion processing in depressed suicide attempters," in *2015 7th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, 2015, pp. 37–40.

[27] M. Lotito and E. Cook, "A review of suicide risk assessment instruments and approaches," *Mental Health Clinician*, vol. 5, no. 5, pp. 216–223, 2015.

[28] Z. Tan, X. Liu, X. Liu, Q. Cheng, and T. Zhu, "Designing microblog direct messages to engage social media users with suicide ideation: interview and survey study on weibo," *Journal of medical Internet research*, vol. 19, no. 12, 2017.

[29] Y.-P. Huang, T. Goh, and C. L. Liew, "Hunting suicide notes in web 2.0-preliminary findings," in *IEEE International Symposium on Multimedia Workshops*. IEEE, 2007, pp. 517–521.

[30] K. D. Varathan and N. Talib, "Suicide detection system based on twitter," in *Science and Information Conference (SAI)*. IEEE, 2014, pp. 785–788.

[31] J. Jashinsky, S. H. Burton, C. L. Hanson, J. West, C. Giraud-Carrier, M. D. Barnes, and T. Argyle, "Tracking suicide risk factors through twitter in the us," *Crisis*, 2014.

[32] J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," *Suicidologi*, vol. 17, no. 3, 2015.

[33] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM*, 2015.

[34] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on twitter," *Computer communications*, vol. 73, pp. 291–300, 2016.

[35] G. Coppersmith, K. Ngo, R. Leary, and A. Wood, "Exploratory analysis of social media prior to a suicide attempt," in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 2016, pp. 106–117.

[36] P. Solano, M. Ustulin, E. Pizzorno, M. Vichi, M. Pompili, G. Serafini, and M. Amore, "A google-based approach for monitoring suicide risk," *Psychiatry research*, vol. 246, pp. 581–586, 2016.

[37] H. Y. Huang and M. Bashir, "Online community and suicide prevention: Investigating the linguistic cues and reply bias," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2016.

[38] M. De Choudhury and E. Kiciman, "The language of social support in social media and its effect on suicidal ideation risk," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[39] M. E. Larsen, N. Cummins, T. W. Boonstra, B. O'Dea, J. Tighe, J. Nicholas, F. Shand, J. Epps, and H. Christensen, "The use of technology in suicide prevention," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 7316–7319.

[40] N. Masuda, I. Kurahashi, and H. Onari, "Suicide ideation of individuals in online social networks," *PloS one*, vol. 8, no. 4, p. e62262, 2013.

[41] S. Chattopadhyay, "A study on suicidal risk analysis," in *International Conference on e-Health Networking, Application and Services*. IEEE, 2007, pp. 74–78.

[42] D. Delgado-Gomez, H. Blasco-Fontecilla, F. Sukno, M. S. Ramos-Plasencia, and E. Baca-Garcia, "Suicide attempters classification: Toward predictive models of suicidal behavior," *Neurocomputing*, vol. 92, pp. 3–8, 2012.

[43] S. Chattopadhyay, "A mathematical model of suicidal-intent-estimation in adults," *American Journal of Biomedical Engineering*, vol. 2, no. 6, pp. 251–262, 2012.

[44] W. Wang, L. Chen, M. Tan, S. Wang, and A. P. Sheth, "Discovering fine-grained sentiment in suicide notes," *Biomedical informatics insights*, vol. 5, no. Suppl 1, p. 137, 2012.

[45] A. Abboute, Y. Boudjeriou, G. Entringer, J. Azé, S. Bringay, and P. Poncelet, "Mining twitter for suicide prevention," in *International Conference on Applications of Natural Language to Data Bases/Information Systems*. Springer, 2014, pp. 250–253.

[46] E. Okhapkina, V. Okhapkin, and O. Kazarin, "Adaptation of information retrieval methods for identifying of destructive informational influence in social networks," in *31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2017, pp. 87–92.

[47] M. Mulholland and J. Quinn, "Suicidal tendencies: The automatic classification of suicidal and non-suicidal lyricists using nlp." in *IJCNLP*, 2013, pp. 680–684.

[48] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, and T. Zhu, "Detecting suicidal ideation in chinese microblogs with psychological lexicons," in *IEEE 11th International Conference on Ubiquitous Intelligence and Computing and Autonomic and Trusted Computing, and IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom)*. IEEE, 2014, pp. 844–849.

[49] X. Huang, X. Li, T. Liu, D. Chiu, T. Zhu, and L. Zhang, "Topic model for identifying suicidal ideation in chinese microblog," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, 2015, pp. 553–562.

[50] Y.-M. Tai and H.-W. Chiu, "Artificial neural network analysis on suicide and self-harm history of taiwanese soldiers," in *Second International Conference on Innovative Computing, Information and Control*. IEEE, 2007, pp. 363–363.

[51] M. Liakata, J. H. Kim, S. Saha, J. Hastings, and D. Rebholzschuhmann, "Three hybrid classifiers for the detection of emotions in suicide notes," *Biomedical Informatics Insights*, vol. 2012, no. (Suppl. 1), pp. 175–184, 2012.

[52] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, and A. Leenaars, "Suicide note classification using natural language processing: A content analysis," *Biomedical informatics insights*, vol. 2010, no. 3, p. 19, 2010.

[53] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating machine learning algorithms for twitter data against established measures of suicidality," *JMIR mental health*, vol. 3, no. 2, p. e21, 2016.

[54] A. L. Nobles, J. J. Glenn, K. Kowsari, B. A. Teachman, and L. E. Barnes, "Identification of imminent suicide risk among young adults using text messages," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–11.

[55] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[56] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[57] S. Ji, G. Long, S. Pan, T. Zhu, J. Jiang, S. Wang, and X. Li, "Knowledge transferring via model aggregation for online social care," *arXiv preprint arXiv:1905.07665*, 2019.

[58] M. Gaur, A. Alambo, J. P. Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, and J. Pathak, "Knowledge-aware assessment of severity of suicide risk for early intervention," in *The World Wide Web Conference*. ACM, 2019, pp. 514–525.

[59] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomedical informatics insights*, 2018.

[60] R. Sawhney, P. Manchanda, P. Mathur, R. Shah, and R. Singh, "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*, 2018, pp. 167–175.

[61] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," *Algorithms*, vol. 13, no. 1, p. 7, 2020.

[62] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," *arXiv preprint arXiv:2004.07601*, 2020.

[63] A. Zirikly, P. Resnik, O. Uzuner, and K. Hollingshead, "Clpsych 2019 shared task: Predicting the degree of suicide risk in reddit posts," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 24–33.

[64] A. G. Hevia, R. C. Menéndez, and D. Gayo-Avello, "Analyzing the use of existing systems for the clpsych 2019 shared task," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 148–151.

[65] M. Morales, P. Dey, T. Theisen, D. Belitz, and N. Chernova, "An investigation of deep learning systems for suicide risk assessment," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 177–181.

[66] M. Matero, A. Idnani, Y. Son, S. Giorgi, H. Vu, M. Zamani, P. Limbachiya, S. C. Guntuku, and H. A. Schwartz, "Suicide risk assessment with multi-level dual-context language and bert," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 39–44.

[67] L. Chen, A. Aldayel, N. Bogoychev, and T. Gong, "Similar minds post alike: Assessment of suicide risk using a hybrid model," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 152–157.

[68] X. Zhao, S. Lin, and Z. Huang, "Text classification of micro-blog's tree hole based on convolutional neural network," in *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*. ACM, 2018, p. 61.

[69] T. Tran, D. Phung, W. Luo, R. Harvey, M. Berk, and S. Venkatesh, "An integrated framework for suicide risk prediction," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1410–1418.

[70] S. Berrouiguet, R. Billot, P. Lenca, P. Tanguy, E. Baca-Garcia, M. Simonnet, and B. Gourvennec, "Toward e-health applications for suicide prevention," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2016, pp. 346–347.

[71] D. Meyer, J.-A. Abbott, I. Rehm, S. Bhar, A. Barak, G. Deng, K. Wallace, E. Ogden, and B. Klein, "Development of a suicidal ideation detection tool for primary healthcare settings: using open access online psychosocial data," *Telemedicine and e-Health*, vol. 23, no. 4, pp. 273–281, 2017.

[72] A. P. Shah, V. Vaibhav, V. Sharma, M. Al Ismail, J. Girard, and L.-P. Morency, "Multimodal behavioral markers exploring suicidal intent in social media videos," in *2019 International Conference on Multimodal Interaction*, 2019, pp. 409–413.

[73] K. M. Harris, J. P. McLean, and J. Sheffield, "Suicidal and online: How do online behaviors inform us of this high-risk population?" *Death studies*, vol. 38, no. 6, pp. 387–394, 2014.

[74] H. Sueki, "The association of suicide-related twitter use with suicidal behaviour: a cross-sectional study of young internet users in japan," *Journal of affective disorders*, vol. 170, pp. 155–160, 2015.

[75] K. W. Hammond, R. J. Laundry, T. M. OLeary, and W. P. Jones, "Use of text search to effectively identify lifetime prevalence of suicide attempts among veterans," in *2013 46th Hawaii International Conference on System Sciences*. IEEE, 2013, pp. 2676–2683.

[76] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting risk of suicide attempts over time through machine learning," *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, 2017.

[77] T. Iliou, G. Konstantopoulou, M. Ntekouli, D. Lymberopoulos, K. Assimakopoulos, D. Galiatsatos, and G. Anastassopoulos, "Machine learning preprocessing method for suicide prediction," in *Artificial Intelligence Applications and Innovations*, L. Iliadis and I. Maglogiannis, Eds. Cham: Springer International Publishing, 2016, pp. 53–60.

[78] T. Nguyen, T. Tran, S. Gopakumar, D. Phung, and S. Venkatesh, "An evaluation of randomized machine learning methods for redundant data: Predicting short and medium-term suicide risk from administrative records and risk assessments," *arXiv preprint arXiv:1605.01116*, 2016.

[79] H. S. Bhat and S. J. Goldman-Mellor, "Predicting adolescent suicide attempts with neural networks," in *NIPS 2017 Workshop on Machine Learning for Health*, 2017.

[80] J. P. Pestian, P. Matykiewicz, M. Linn-Gust, B. South, O. Uzuner, J. Wiebe, K. B. Cohen, J. Hurdle, and C. Brew, "Sentiment analysis of suicide notes: A shared task," *Biomedical informatics insights*, vol. 5, no. Suppl. 1, p. 3, 2012.

[81] E. White and L. J. Mazlack, "Discerning suicide notes causality using fuzzy cognitive maps," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ)*. IEEE, 2011, pp. 2940–2947.

[82] B. Desmet and V. Hoste, "Emotion detection in suicide notes," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6351–6358, 2013.

[83] R. Wicentowski and M. R. Sydes, "emotion detection in suicide notes using maximum entropy classification," *Biomedical informatics insights*, vol. 5, pp. BII–S8972, 2012.

[84] A. Kovačević, A. Dehghan, J. A. Keane, and G. Nenadic, "Topic categorisation of statements in suicide notes with integrated rules and machine learning," *Biomedical informatics insights*, vol. 5, pp. BII–S8978, 2012.

[85] A. M. Schoene and N. Dethlefs, "Automatic identification of suicide notes from linguistic and sentiment features," in *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 2016, pp. 128–133.

[86] J. Robinson, G. Cox, E. Bailey, S. Hetrick, M. Rodrigues, S. Fisher, and H. Herrman, "Social media and suicide prevention: a systematic review," *Early intervention in psychiatry*, vol. 10, no. 2, pp. 103–121, 2016.

[87] Y. Wang, S. Wan, and C. Paris, "The role of features and context on suicide ideation detection," in *Proceedings of the Australasian Language Technology Association Workshop 2016*, 2016, pp. 94–102.

[88] A. Shepherd, C. Sanders, M. Doyle, and J. Shaw, "Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation," *BMC psychiatry*, vol. 15, no. 1, p. 29, 2015.

[89] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity." in *ICWSM*, 2014.

[90] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "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*. ACM, 2016, pp. 2098–2110.

[91] M. Kumar, M. Dredze, G. Coppersmith, and M. De Choudhury, "Detecting changes in suicide content manifested in social media following celebrity suicides," in *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 2015, pp. 85–94.

[92] L. Guan, B. Hao, Q. Cheng, P. S. Yip, and T. Zhu, "Identifying chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model," *JMIR mental health*, vol. 2, no. 2, p. e17, 2015.

[93] S. J. Cash, M. Thelwall, S. N. Peck, J. Z. Ferrell, and J. A. Bridge, "Adolescent suicide statements on myspace," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 3, pp. 166–174, 2013.

[94] I. Gilat, Y. Tobin, and G. Shahar, "Offering support to suicidal individuals in an online support group," *Archives of Suicide Research*, vol. 15, no. 3, pp. 195–206, 2011.

[95] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2968–2978.

[96] Y. Wang, J. Tang, J. Li, B. Li, Y. Wan, C. Mellina, N. O'Hare, and Y. Chang, "Understanding and discovering deliberate self-harm content in social media," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 93–102.

[97] Q. Li, Y. Xue, L. Zhao, J. Jia, and L. Feng, "Analyzing and identifying teens stressful periods and stressor events from a microblog," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1434–1448, 2016.

[98] Z. Huang, J. Yang, F. van Harmelen, and Q. Hu, "Constructing knowledge graphs of depression," in *International Conference on Health Information Science*. Springer, 2017, pp. 149–161.

[99] F. Hao, G. Pang, Y. Wu, Z. Pi, L. Xia, and G. Min, "Providing appropriate social support to prevention of depression for highly anxious sufferers," *IEEE Transactions on Computational Social Systems*, 2019.

[100] A. E. Aladağ, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu, and H. O. Bingol, "Detecting suicidal ideation on forums: proof-of-concept study," *Journal of medical Internet research*, vol. 20, no. 6, p. e215, 2018.

[101] D. N. Milne, G. Pink, B. Hachey, and R. A. Calvo, "CLPsych 2016 shared task: Triaging content in online peer-support forums," in *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, 2016, pp. 118–127.

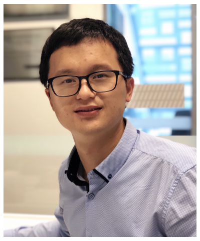
[102] K. Haerian, H. Salmasian, and C. Friedman, "Methods for identifying suicide or suicidal ideation in ehrs," in *AMIA annual symposium proceedings*, vol. 2012. American Medical Informatics Association, 2012, p. 1244.

[103] D. E. Losada and F. Crestani, "A test collection for research on depression and language use," in *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 2016, pp. 28–39.

[104] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "Clpsych 2015 shared task: Depression and ptsd on twitter," in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 31–39.

[105] X. Liu, X. Liu, J. Sun, N. X. Yu, B. Sun, Q. Li, and T. Zhu, "Proactive suicide prevention online (pspo): Machine identification and crisis management for chinese social media users with suicidal thoughts and behaviors," *Journal of Medical Internet Research*, vol. 21, no. 5, p. e11705, 2019.

Shaoxiong Ji is a doctoral candidate at the Department of Computer Science, Aalto University, Finland. He received his bachelor's degree from Dalian University of Technology, China. He worked as a research assistant or visiting researcher at the University of Technology Sydney, The University of Queensland, and Nanyang Technological University. His research interests include machine learning and data mining.

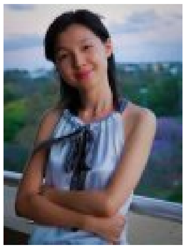


Shirui Pan received his Ph.D. degree in computer science from University of Technology Sydney (UTS), Australia, in 2015. He is a Lecturer with the Faculty of IT at Monash University. Since 2010, he has published over 80 research papers in top-tier journals and conferences, including IEEE Transactions on Neural Networks and Learning Systems (TNNLS), IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Transactions on Cybernetics (TCYB), Pattern Recognition, International Joint Conference on Artificial Intelligence (IJCAI), International Conference on Data Engineering (ICDE) and IEEE International Conference on Data Mining (ICDM). His current research interests include data mining, machine learning, and graph data analytics.

Xue Li received the Ph.D. degree from the Queensland University of Technology, Brisbane, QLD, Australia, in 1997. He is a Professor with the School of ITEE, University of Queensland, St Lucia, QLD, Australia. His current research interests include data mining, social computing, database systems, and intelligent Web information systems.

Erik Cambria is the Founder of SenticNet, a Singapore-based company offering B2B sentiment analysis services, and an Associate Professor at NTU, where he also holds the appointment of Provost Chair in Computer Science and Engineering. Prior to joining NTU, he worked at Microsoft Research Asia and HP Labs India and earned his PhD through a joint programme between the University of Stirling and MIT Media Lab. He is Associate Editor of several journals, e.g., NEUCOM, INFFUS, KBS, IEEE CIM and IEEE Intelligent Systems (where he manages the Department of Affective Computing and Sentiment Analysis), and is involved in many international conferences as PC member and program chair.

Guodong Long received his PhD degree from the University of Technology Sydney (UTS), Australia, in 2014. He is a senior lecturer at the Australian Artificial Intelligence Institute, Faculty of Engineering and IT, UTS. His research focuses on data mining, machine learning, and NLP. He has more than 40 research papers published on top-tier journals, including IEEE TPAMI, TCYB, and TKDE, and conferences including ICLR, AAAI, IJCAI, and ICDM.



Zi Huang received the B.Sc. degree in computer science from Tsinghua University, China, and the Ph.D. degree in computer science from the School of Information Technology and Electrical Engineering, The University of Queensland, Australia. She is an ARC Future Fellow with the School of Information Technology and Electrical Engineering, The University of Queensland. Her research interests include multimedia indexing and search, social data analysis, and knowledge discovery.