

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

Shaoxiong Ji¹, Shirui Pan², *Member, IEEE*, Xue Li, *Member, IEEE*, Erik Cambria³, *Senior Member, IEEE*, Guodong Long⁴, and Zi Huang⁵, *Member, IEEE*

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 (SID) 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 article is the first survey that comprehensively introduces and discusses the methods from these categories. Domain-specific applications of SID are reviewed according to their data sources, i.e., questionnaires, electronic health records, suicide notes, and online user content. Several specific tasks and data sets 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—Deep learning, feature engineering, social content, suicidal ideation detection (SID).

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.

Manuscript received October 17, 2019; revised April 26, 2020, June 22, 2020, and August 16, 2020; accepted August 31, 2020. (*Corresponding author: Shaoxiong Ji.*)

Shaoxiong Ji is with the Department of Computer Science, Aalto University, 02150 Espoo, Finland, and also with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: shaoxiong.ji@aalto.fi).

Shirui Pan is with the Faculty of Information Technology, Monash University, Melbourne, VIC 3800, Australia (e-mail: shirui.pan@monash.edu).

Xue Li and Zi Huang are with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: xueli@itee.uq.edu.au; huang@itee.uq.edu.au).

Erik Cambria is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: cambria@ntu.edu.sg).

Guodong Long is with the Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: guodong.long@uts.edu.au).

Digital Object Identifier 10.1109/TCSS.2020.3021467

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 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. SID 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

¹<https://thesun.co.uk/news/worldnews/3003805>

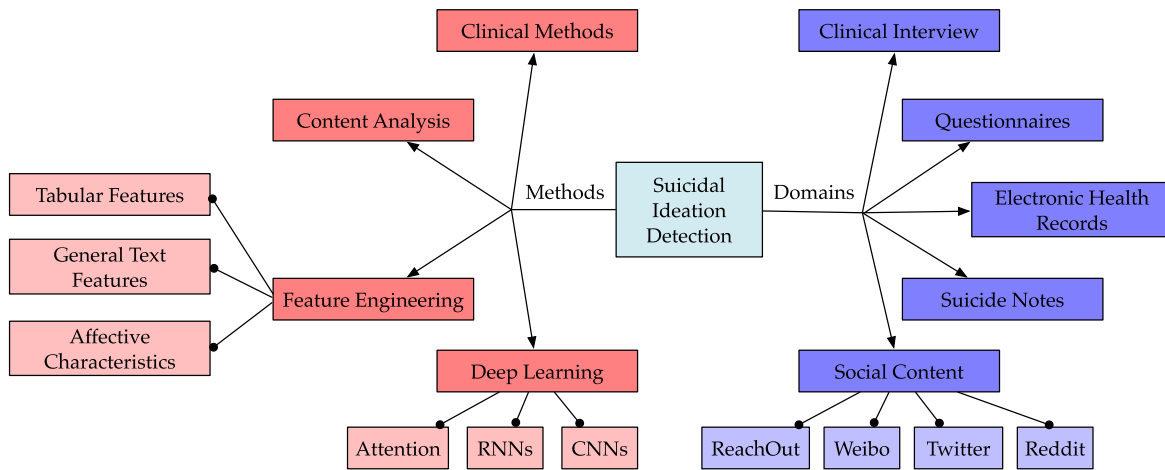


Fig. 1. Categorization of suicide ideation detection: methods and domains. The left part represents method categorization, while the right part shows the categories of domains. The arrow and solid point indicate subcategories.

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]–[18]. Those methods generally require heuristics to select features or design artificial neural network (ANN) 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 Many other suicide prevention tools integrated with social networking services have also been developed, including Samaritans Radar³ 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 SID 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 are a limited number of benchmarks for training and evaluating SID. AI-powered models, sometimes, learn statistical clues but fail to understand people's intentions. Moreover, many neural models are lack of interpretability. This survey reviews SID 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 article provides a comprehensive review of the increasingly important field of SID 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.

- 1) To the best of our knowledge, this is the first survey that conducts a comprehensive review of SID, its methods, and its applications from a machine learning perspective.
- 2) 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.
- 3) We enumerate existing and less explored tasks and discuss their limitations. We also summarize existing data sets and provide an outlook of future research directions in this field.

The remainder of this article is organized as follows. Methods and applications are introduced and summarized in Sections II and III, respectively. Section IV enumerates specific tasks and some data sets. 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.

²<https://blackdoginstitute.org.au/research/digital-dog/programs/ibobbly-app>

³<https://samaritans.org/about-samaritans/research-policy/internet-suicide/samaritans-radar>

⁴<https://woebot.io>

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 nonsuicidal juveniles. Other clinical methods examine the resting state heart rate from converted sensing signals [24] and classify the 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, we 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 pointwise mutual information symptom lexicon using an annotated Twitter data set. Gunn and Lester [32] analyzed posts from Twitter in the 24 h 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, hopelessness, or other social factors, such as 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.

1) *Tabular Features*: Tabular data for SID 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 (SRRS). Chattopadhyay [43] proposed to apply a multilayer feedforward neural network, as shown in Fig. 2(a), 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 machine (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] extracted 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. 2(b), 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, LIWC, emotion features, and mental disease lexicon.

Models for SID with feature engineering include SVM [44], ANNs [50], and conditional random field (CRF) [51]. Tai and Chiu [50] selected several features, including the history of suicide ideation and self-harm behavior, religious belief, family status, mental disorder history of candidates, and

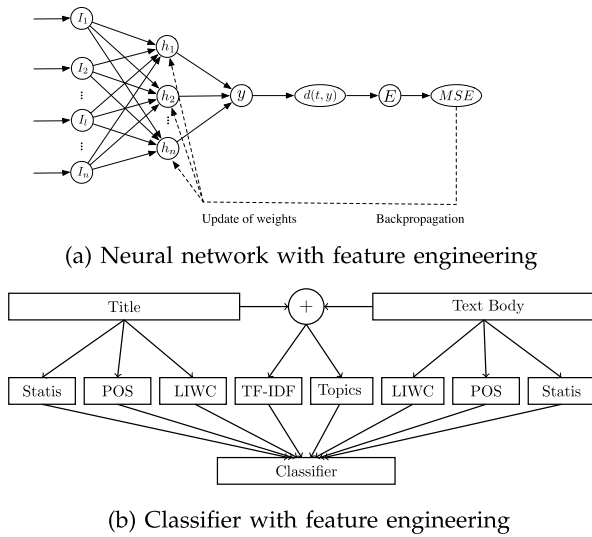


Fig. 2. Illustrations of methods with feature engineering. (a) Neural network with feature engineering. (b) Classifier with feature engineering.

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 that machine learning algorithms can effectively identify high suicidal risk.

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 (two 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. Especially, 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 SID and suicide prevention. It can effectively learn text features automatically without

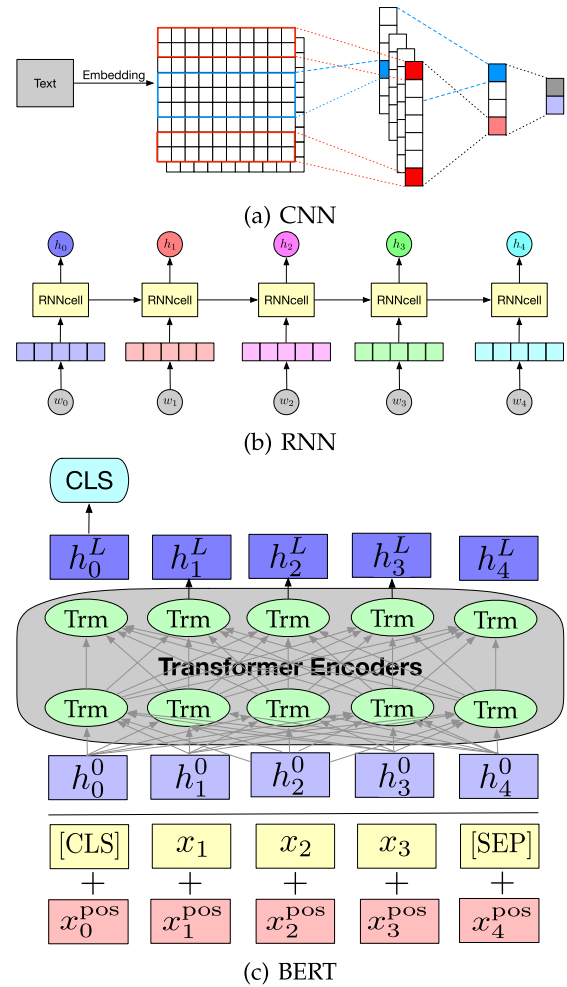


Fig. 3. DNNs for SID. (a) CNN. (b) RNN. (c) BERT.

sophisticated feature engineering techniques. At the same time, some also take extracted features into deep neural networks (DNNs); for example, Nobles *et al.* [54] fed psycholinguistic features and word occurrence into the multilayer perceptron (MLP). Popular DNNs include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT), as shown in Fig. 3(a)–(c). 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. The 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 SID. 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 semisupervised learning methods. Benton *et al.* [16] predicted

suicide attempt and mental health with neural models under the framework of multitask 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 SID. Similarly, Tadesse *et al.* [61] employed the 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 a dual-context model using hierarchically attentive RNN and BERT.

Another subdirection is the so-called hybrid method that 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.

D. Summary

The popularization of machine learning has facilitated research on SID from multimodal 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, readability, and deep learning models, such as CNN and LSTM. Those methods, especially DNNs 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 toward knowledge-aware detection.

III. APPLICATIONS ON DOMAINS

Many machine learning techniques have been introduced for SID. 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 SID, where Fig. 4(a) lists selected questions of the “International Personal Disorder Examination Screening Questionnaire” (IPDE-SQ) adapted from [42], Fig. 4(b) shows selected patient’s records

from [69], Fig. 4(c) shows a suicide note from a website,⁵ and Fig. 4(d) shows 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. Berrouguet *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 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 four-point ratings, while the IPDE-SQ in DSM-IV has 77 true-false screening questions. Furthermore, Delgado-Gomez *et al.* [42] introduced the “Holmes–Rahe SRRS” and the IPDE-SQ as well to two comparison groups of suicide attempters and nonsuicide 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 EHRs has paved the way for machine learning techniques for suicide attempter prediction. Patient records include demographical information and diagnosis-related history, such as 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

⁵<https://paranorms.com/suicide-notes>

⁶<https://psychiatry.org/psychiatrists/practice/dsm>

⁷<https://apps.who.int/classifications/icd10/browse/2016/en>

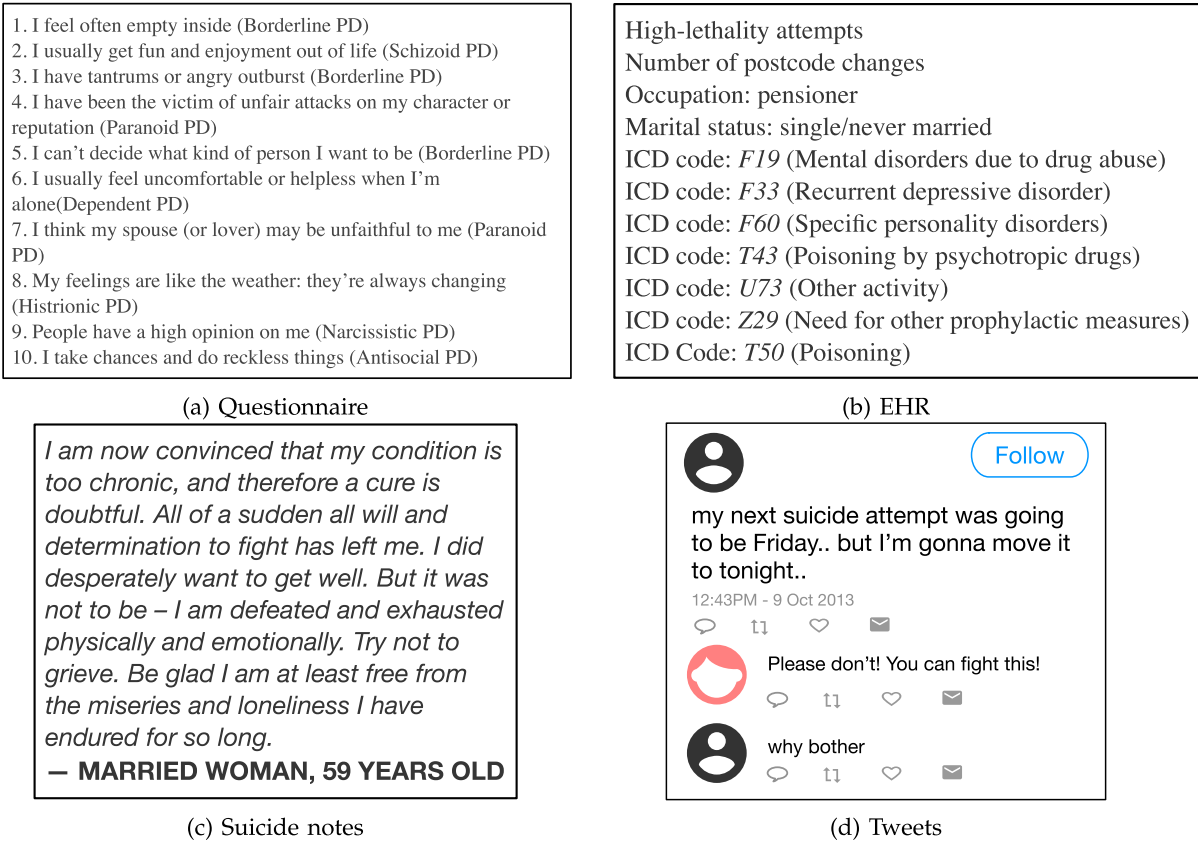


Fig. 4. Examples of content for SID. (a) Questionnaire. (b) EHR. (c) Suicide notes. (d) Tweets.

TABLE I
CATEGORIZATION OF METHODS FOR SID

Category	Publications	Methods	Inputs
Feature Engineering	Ji <i>et al.</i> , 2018 [17]	Word counts, POS, LIWC, TF-IDF + classifiers	Text
	Masuda <i>et al.</i> , 2013 [40]	Multivariate/univariate logistic regression	Characteristics variables
Deep Learning	Delgado-Gomez <i>et al.</i> , 2012 [42]	International personal disorder examination screening questionnaire	Questionnaire responses
	Mulholland and Quinn, 2013 [47]	Vocabulary features, syntactic features, semantic class features, N-gram	lyrics
	Okhapkina <i>et al.</i> , 2017 [46]	Dictionary, TF-IDF + SVD	Text
	Huang <i>et al.</i> , 2014 [48]	Lexicon, syntactic features, POS, tense	Text
	Pestian <i>et al.</i> , 2010 [52]	Word counts, POS, concepts and readability score	Text
	Tai and Chiu, 2007 [50]	self-measurement scale + ANN	Self-measurement forms
	Shing <i>et al.</i> , 2018 [13]	BoWs, empath, readability, syntactic, topic, LIWC, emotion, lexicon	Text
	Zhao <i>et al.</i> , 2018 [68]	Word embedding, tabular features, D-CNN	Text+external information
	Shing <i>et al.</i> , 2018 [13]	Word embedding, CNN, max pooling	Text
	Ji <i>et al.</i> , 2018 [17]	Word embedding, LSTM, max pooling	Text
Deep Learning	Benton <i>et al.</i> , 2017 [16]	Multi-task learning, neural networks	Text
	Nobles <i>et al.</i> , 2018 [54]	MLP, psycholinguistic features, word occurrence	Text
	Hevia <i>et al.</i> , 2019 [64]	Pretrained GRU, word embedding, document embedding	Text
	Morales <i>et al.</i> , 2019 [65]	CNN, LSTM, NeuNetS, word embedding	Text
	Matero <i>et al.</i> , 2019 [66]	Dual-context, BERT, GRU, attention, user-factor adaptation	Text
	Gaur <i>et al.</i> , 2019 [58]	CNN, knowledge base, ConceptNet embedding	Text
	Coppersmith <i>et al.</i> , 2018 [59]	GloVe, BiLSTM, self attention	Text
	Ji <i>et al.</i> , 2020 [62]	Relation network, LSTM, attention, lexicon	Text
	Tadesse <i>et al.</i> , 2020 [61]	LSTM, CNN, word embedding	Text
			Text

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 data set to relax the restriction on patient's diagnosis-related history and applied neural networks as a classification model to predict suicide attempters.

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 [44], [80], 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 multiclass annotations.

Emotion in text provides sentimental cues of suicidal ideation understanding. Desmet and Hoste [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 identify 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, such as 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 [17], [90]. 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 SID. 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- 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] used the blog stream collected from public blog articles written by suicide victims to study the accumulated emotional information.

E. Summary

Applications of SID mainly consist of four domains: questionnaires, EHRs, 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 SID. 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.

IV. TASKS AND DATA SETS

In this section, we summarize specific tasks in SID 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, 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

TABLE II
SUMMARY OF STUDIES ON SID 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]–[26] mobile applications [19] content analysis II-A feature engineering II-B deep learning II-C
Critical issues	suicide factors [1]–[5] ethics [20]–[22] privacy [20]

list of data sets 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 multiclass classification. Binary suicidality classification simply determines text with suicidal ideation or not, while multiclass suicidality classification conducts fine-grained suicide risk assessment. For example, some studies divide suicide risk into four levels: no, low, moderate, and severe. Alternatively, it can also consider four types of class labels according to mental and behavioral procedures, i.e., nonsuicidal, 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 multiaspect labels, including mental pain, cognitive attribution, and level of suicidal risk. Mental pain includes the 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, finding an explanation of suicidal risk in

mental pain, and cognitive attribution aspects associated with suicide.

3) *Suicide Attempter Detection*: The two tasks mentioned earlier focus on a single text itself. However, the primary purpose of SID is the identify suicide attempters. Thus, it is vital to achieving user-level detection, which consists of two folds, i.e., user-level multiinstance suicidality detection and suicide attempt detection on a graph. The former takes a bag of posts from individuals as input and conducts multiinstance 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 SID 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. Data Sets

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 data set with 3549 posts with suicidal ideation. Shing *et al.* [13] published their UMD Reddit Suicidality data set 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,

⁸<https://reddit.com/r/SuicideWatch>

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 data set of 594 tweets with suicidal ideation out of a total of 10288 tweets. Vioulès *et al.* collected 5446 tweets using Twitter streaming API [3], of which 2381 and 3065 tweets are from the distressed users and normal users, respectively. However, most Twitter-based data sets are no longer available as per the policy of Twitter.

c) *ReachOut*: ReachOut Forum⁹ is a peer support platform provided by an Australian mental health care organization. The ReachOut data set [101] was first released in the CLPsych17 shared task. Participants were initially given a training data set of 65756 forum posts, of which 1188 were annotated manually with the expected category, and a test set of 92207 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 EHRs is from the California emergency department encounter and hospital admission. It contains 522056 anonymous EHR records from California-resident adolescents. However, it is not public for access. Bhat and Goldman-Mellor [79] first used these records from 2006 to 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 the 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 data set contains 7746 patients and 17771 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 data set 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 data set contains sequential text from social media. Another data set is the Reddit Self-reported Depression Diagnosis (RSDD) data set [95], which contains 9000 diagnosed users with depression and approximately 107000 matched control users.

V. DISCUSSION AND FUTURE WORK

Many preliminary works have been conducted for SID, 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

1) *Data Deficiency*: The most critical issue of current research is data deficiency. Current methods mainly apply supervised learning techniques that 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 multispect data and data with social relationships.

2) *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.

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

4) *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 SID. 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 SID, and generative adversarial networks can be used to generate 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.

⁹<https://au.reachout.com/forums>

¹⁰<https://early.irlab.org>

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

TABLE III
SUMMARY OF THE PUBLIC DATA SETS

Type	Publication	Source	Instances	Public Access
Text	Shing <i>et al.</i> , 2018 [13]	Reddit	866/11,129	https://tinyurl.com/umnd-suicidality
Text	Aladağ <i>et al.</i> , 2018 [100]	Reddit	508,398	Request to the authors
Text	Coppersmith <i>et al.</i> , 2015b [33]	Twitter	> 3,200	N.A
Text	Coppersmith <i>et al.</i> , 2015a [104]	Twitter	1,711	https://tinyurl.com/clpsych-2015
Text	Vioulès <i>et al.</i> , 2018 [3]	Twitter	5,446	N.A
Text	Milne <i>et al.</i> , 2016 [101]	ReachOut	65,756	N.A
EHR	Bhat and Goldman-Mellor, 2017 [79]	Hospital	522,056	N.A
EHR	Tran <i>et al.</i> , 2013 [69]	Barwon Health	7,746	N.A
EHR	Haerian <i>et al.</i> , 2012 [102]	CDW & WebCIS	280	N.A
Text	Pestian <i>et al.</i> , 2012 [80]	Notes	1,319	2011 i2b2 NLP challenge
Text	Gaur <i>et al.</i> , 2019 [58]	Reddit	500/2,181	Request to the authors
Text	Losada and Crestani, 2016 [103]	Social networks	892	https://tec.citius.usc.es/ir/code/eRisk.html
Text	Yates <i>et al.</i> , 2017 [95]	Reddit	9,000	http://ir.cs.georgetown.edu/resources/rsdd.html

2) *Suicidal Intention Understanding and Interpretability*: Many factors are correlated with 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 SID 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 SID from a broad perspective that covers clinical methods, such as 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, such as 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 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 SID 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

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

[1] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide," *Arch. Suicide Res.*, vol. 14, no. 3, pp. 206–221, Jul. 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*, vol. 37, no. 2, pp. 161–165, Mar. 2016.
- [3] M. J. Vioules, B. Moulahi, J. Aze, and S. Bringay, "Detection of suicide-related posts in Twitter data streams," *IBM J. Res. Develop.*, vol. 62, no. 1, pp. 7:1–7:12, Jan. 2018.
- [4] A. J. Ferrari *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, Apr. 2014, Art. no. e91936.
- [5] R. C. O'Connor and M. K. Nock, "The psychology of suicidal behaviour," *Lancet Psychiatry*, vol. 1, no. 1, pp. 73–85, 2014.
- [6] J. Lopez-Castroman *et al.*, "Mining social networks to improve suicide prevention: A scoping review," *J. Neurosci. Res.*, vol. 98, no. 4, pp. 616–625, Apr. 2020.
- [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, Mar. 2019.
- [8] G. Kassen, A. Kudaibergenova, A. Mukasheva, D. Yertargynkyzy, and K. Moldassan, "Behavioral risk factors for suicide among adolescent schoolchildren," *Elementary Educ. Online*, vol. 19, pp. 66–77, Jan. 2020.
- [9] V. Venek, S. Scherer, L.-P. Morency, A. S. Rizzo, and J. Pestian, "Adolescent suicidal risk assessment in clinician-patient interaction," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 204–215, Apr. 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," *Artif. Intell. Med.*, vol. 52, no. 3, pp. 165–168, Jul. 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 Trans. Comput. Social Syst.*, vol. 6, no. 5, pp. 907–921, Oct. 2019.
- [12] B. O'Dea, S. Wan, P. J. Batterham, A. L. Cleave, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Intervent.*, vol. 2, no. 2, pp. 183–188, May 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 *Proc. 5th Workshop Comput. Linguistics Clin. Psychol., Keyboard 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 J. Biomed. Health Informat.*, vol. 20, no. 5, pp. 1384–1396, Sep. 2016.
- [15] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, pp. 1–47, Aug. 2018.
- [16] A. Benton, M. Mitchell, and D. Hovy, "Multi-task learning for mental health using social media text," in *Proc. EACL*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, pp. 152–162.
- [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*, vol. 2018, pp. 1–10, Sep. 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 *Proc. 24th Int. Conf. Database Syst. Adv. Appl. (DASFAA)*, Cham, Switzerland: Springer, 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, Jan. 2017, Art. no. e013518.
- [20] N. N. Gomes de Andrade, D. Pawson, D. Muriello, L. Donahue, and J. Guadagno, "Ethics and artificial intelligence: Suicide prevention on Facebook," *Philosophy Technol.*, vol. 31, no. 4, pp. 669–684, Dec. 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 Psychiatry*, vol. 9, p. 650, Dec. 2018.
- [22] K. P. Linthicum, K. M. Schafer, and J. D. Ribeiro, "Machine learning in suicide science: Applications and ethics," *Behav. Sci. Law*, vol. 37, no. 3, pp. 214–222, May 2019.
- [23] S. Scherer, J. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 709–713.
- [24] D. Sikander *et al.*, "Predicting risk of suicide using resting state heart rate," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–4.
- [25] M. A. Just *et al.*, "Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth," *Nature Hum. Behav.*, vol. 1, no. 12, pp. 911–919, 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 *Proc. 7th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Nov. 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, Sep. 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," *J. Med. Internet Res.*, vol. 19, no. 12, p. e381, Dec. 2017.
- [29] Y.-P. Huang, T. Goh, and C. L. Liew, "Hunting suicide notes in Web 2.0—preliminary findings," in *Proc. 9th IEEE Int. Symp. Multimedia Workshops (ISMW)*, Dec. 2007, pp. 517–521.
- [30] K. D. Varathan and N. Talib, "Suicide detection system based on Twitter," in *Proc. Sci. Inf. Conf.*, Aug. 2014, pp. 785–788.
- [31] J. Jashinsky *et al.*, "Tracking suicide risk factors through Twitter in the US," *Crisis*, vol. 35, no. 1, pp. 51–59, Jan. 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, pp. 28–30, Jun. 2015.
- [33] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in *Proc. Joint Statist. Meetings Stat. Comput. Sect. (JSM)*, 2015, pp. 1–15.
- [34] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on Twitter," *Comput. Commun.*, vol. 73, pp. 291–300, Jan. 2016.
- [35] G. Coppersmith, K. Ngo, R. Leary, and A. Wood, "Exploratory analysis of social media prior to a suicide attempt," in *Proc. 3rd Workshop Comput. Linguistics Clin. Psychol.*, 2016, pp. 106–117.
- [36] P. Solano *et al.*, "A Google-based approach for monitoring suicide risk," *Psychiatry Res.*, vol. 246, pp. 581–586, Dec. 2016.
- [37] H. Y. Huang and M. Bashir, "Online community and suicide prevention: Investigating the linguistic cues and reply bias," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 1–5.
- [38] M. De Choudhury and E. Kiciman, "The language of social support in social media and its effect on suicidal ideation risk," in *Proc. 11th Int. AAAI Conf. Web Social Media*, 2017, p. 32.
- [39] M. E. Larsen *et al.*, "The use of technology in suicide prevention," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 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, Apr. 2013, Art. no. e62262.
- [41] S. Chattopadhyay, "A study on suicidal risk analysis," in *Proc. 9th Int. Conf. e-Health Netw., Appl. Services*, Jun. 2007, pp. 74–78.
- [42] D. Delgado-Gomez, H. Blasco-Fontecilla, F. Sukno, M. Socorro Ramos-Plasencia, and E. Baca-Garcia, "Suicide attempters classification: Toward predictive models of suicidal behavior," *Neurocomputing*, vol. 92, pp. 3–8, Sep. 2012.
- [43] S. Chattopadhyay, "A mathematical model of suicidal-intent-estimation in adults," *Amer. J. Biomed. Eng.*, vol. 2, no. 6, pp. 251–262, Jan. 2013.
- [44] W. Wang, L. Chen, M. Tan, S. Wang, and A. P. Sheth, "Discovering fine-grained sentiment in suicide notes," *Biomed. Informat. Insights*, vol. 5, no. 1, p. 137, 2012.
- [45] A. Abboute, Y. Boudjeriou, G. Entringer, J. Azé, S. Bringay, and P. Poncelet, "Mining Twitter for suicide prevention," in *Proc. Int. Conf. Appl. Natural Lang. Data Bases/Inf. Syst.*, Cham, Switzerland: 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 *Proc. 31st Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, Mar. 2017, pp. 87–92.
- [47] M. Mulholland and J. Quinn, "Suicidal tendencies: The automatic classification of suicidal and non-suicidal lyricists using NLP," in *Proc. 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 *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput. Autonomic Trusted Comput. IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 844–849.

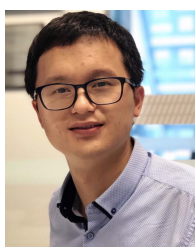
- [49] X. Huang, X. Li, T. Liu, D. Chiu, T. Zhu, and L. Zhang, "Topic model for identifying suicidal ideation in chinese microblog," in *Proc. 29th Pacific Asia Conf. Lang., Inf. Comput.*, 2015, pp. 553–562.
- [50] Y.-M. Tai and H.-W. Chiu, "Artificial neural network analysis on suicide and self-harm history of taiwanese soldiers," in *Proc. 2nd Int. Conf. Innov. Comput., Inf. Control (ICICIC)*, Sep. 2007, p. 363.
- [51] M. Liakata, J. H. Kim, S. Saha, J. Hastings, and D. Rebholz-Schuhmann, "Three hybrid classifiers for the detection of emotions in suicide notes," *Biomed. Informat. Insights*, vol. 2012, no. 1, pp. 175–184, 2012.
- [52] J. Pestian, H. Nasrallah, P. Matykievicz, A. Bennett, and A. Leenaars, "Suicide note classification using natural language processing: A content analysis," *Biomed. Inform. 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, May 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 *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2018, pp. 1–11.
- [55] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: <http://arxiv.org/abs/1301.3781>
- [56] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
- [57] S. Ji *et al.*, "Knowledge transferring via model aggregation for online social care," 2019, *arXiv:1905.07665*. [Online]. Available: <http://arxiv.org/abs/1905.07665>
- [58] M. Gaur *et al.*, "Knowledge-aware assessment of severity of suicide risk for early intervention," in *Proc. World Wide Web Conf. (WWW)*, 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," *Biomed. Informat. Insights*, vol. 10, Jan. 2018, Art. no. 117822261879286.
- [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 *Proc. 9th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 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, Dec. 2019.
- [62] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," 2020, *arXiv:2004.07601*. [Online]. Available: <http://arxiv.org/abs/2004.07601>
- [63] A. Zirikly, P. Resnik, O. Uzunur, and K. Hollingshead, "Clpsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts," in *Proc. 6th Workshop Comput. Linguistics Clin. Psychol.*, 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 *Proc. 6th Workshop Comput. Linguistics Clin. Psychol.*, 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 *Proc. 6th Workshop Comput. Linguistics Clin. Psychol.*, 2019, pp. 177–181.
- [66] M. Matero *et al.*, "Suicide risk assessment with multi-level dual-context language and," in *Proc. 6th Workshop Comput. Linguistics Clin. Psychol.*, 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 *Proc. 6th Workshop Comput. Linguistics Clin. Psychol.*, 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 *Proc. Int. Conf. Algorithms, Comput. Artif. Intell. (ACAI)*, 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 *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2013, pp. 1410–1418.
- [70] S. Berrouguet *et al.*, "Toward E-Health applications for suicide prevention," in *Proc. IEEE 1st Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE)*, Jun. 2016, pp. 346–347.
- [71] D. Meyer *et al.*, "Development of a suicidal ideation detection tool for primary healthcare settings: Using open access online psychosocial data," *Telemedicine e-Health*, vol. 23, no. 4, pp. 273–281, Apr. 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 *Proc. Int. Conf. Multimodal Interact.*, Oct. 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 Stud.*, vol. 38, no. 6, pp. 387–394, Jul. 2014.
- [74] H. Sueki, "The association of suicide-related Twitter use with suicidal behaviour: A cross-sectional study of young Internet users in Japan," *J. Affect. Disorders*, vol. 170, pp. 155–160, Jan. 2015.
- [75] K. W. Hammond, R. J. Laundry, T. M. O'leary, and W. P. Jones, "Use of text search to effectively identify lifetime prevalence of suicide attempts among veterans," in *Proc. 46th Hawaii Int. Conf. Syst. Sci.*, Jan. 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," *Clin. Psychol. Sci.*, vol. 5, no. 3, pp. 457–469, May 2017.
- [77] T. Iliou *et al.*, "Machine learning preprocessing method for suicide prediction," in *Artificial Intelligence Applications and Innovations*, L. Iliadis and I. Maglogiannis, Eds. Cham, Switzerland: Springer, 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," 2016, *arXiv:1605.01116*. [Online]. Available: <http://arxiv.org/abs/1605.01116>
- [79] H. S. Bhat and S. J. Goldman-Mellor, "Predicting adolescent suicide attempts with neural networks," in *Proc. NIPS Workshop Mach. Learn. Health*, 2017, pp. 1–8.
- [80] J. P. Pestian *et al.*, "Sentiment analysis of suicide notes: A shared task," *Biomed. Informat. Insights*, vol. 5s1, Jan. 2012, Art. no. BII.S9042.
- [81] E. White and L. J. Mazlack, "Discerning suicide notes causality using fuzzy cognitive maps," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2011, pp. 2940–2947.
- [82] B. Desmet and V. Hoste, "Emotion detection in suicide notes," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6351–6358, Nov. 2013.
- [83] R. Wicentowski and M. R. Sydes, "Emotion detection in suicide notes using maximum entropy classification," *Biomed. Informat. Insights*, vol. 5s1, Jan. 2012, Art. no. BII.S8972.
- [84] A. Kovačević, A. Dehghan, J. A. Keane, and G. Nenadic, "Topic categorisation of statements in suicide notes with integrated rules and machine learning," *Biomed. Informat. Insights*, vol. 5s1, Jan. 2012, Art. no. BII.S8978.
- [85] A. M. Schoene and N. Dethlefs, "Automatic identification of suicide notes from linguistic and sentiment features," in *Proc. 10th SIGHUM Workshop Lang. Technol. Cultural Heritage, Social Sci., Humanities*, 2016, pp. 128–133.
- [86] J. Robinson *et al.*, "Social media and suicide prevention: A systematic review," *Early Intervent Psychiatry*, vol. 10, no. 2, pp. 103–121, Apr. 2016.
- [87] Y. Wang, S. Wan, and C. Paris, "The role of features and context on suicide ideation detection," in *Proc. Australas. Lang. Technol. Assoc. Workshop*, 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, Dec. 2015.
- [89] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," in *Proc. ICWSM*, 2014, pp. 71–80.
- [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 *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 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 *Proc. 26th ACM Conf. Hypertext Social Media (HT)*, 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, May 2015.
- [93] S. J. Cash, M. Thelwall, S. N. Peck, J. Z. Ferrell, and J. A. Bridge, "Adolescent suicide statements on MySpace," *Cyberpsychology, Behav., Social Netw.*, vol. 16, no. 3, pp. 166–174, Mar. 2013.

- [94] I. Gilat, Y. Tobin, and G. Shahrar, "Offering support to suicidal individuals in an online support group," *Arch. Suicide Res.*, vol. 15, no. 3, pp. 195–206, Jul. 2011.
- [95] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2968–2978.
- [96] Y. Wang et al., "Understanding and discovering deliberate self-harm content in social media," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 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 J. Biomed. Health Informat.*, vol. 21, no. 5, pp. 1434–1448, Sep. 2017.
- [98] Z. Huang, J. Yang, F. van Harmelen, and Q. Hu, "Constructing knowledge graphs of depression," in *Proc. Int. Conf. Health Inf. Sci. Cham*, Switzerland: 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 Trans. Comput. Social Syst.*, vol. 6, no. 5, pp. 879–887, Oct. 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," *J. Med. Internet Res.*, vol. 20, no. 6, p. e215, Jun. 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 *Proc. 3rd Workshop Comput. Linguistics Clin. Psychol.*, 2016, pp. 118–127.
- [102] K. Haerian, H. Salmasian, and C. Friedman, "Methods for identifying suicide or suicidal ideation in EHRs," in *Proc. AMIA Annu. Symp.*, 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 *Proc. Int. Conf. Cross-Lang. Eval. Forum Eur. Lang. Cham*, Switzerland: 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 *Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., Linguistic Signal Clin. Reality*, 2015, pp. 31–39.
- [105] X. Liu et al., "Proactive suicide prevention online (PSPO): Machine identification and crisis management for chinese social media users with suicidal thoughts and behaviors," *J. Med. Internet Res.*, vol. 21, no. 5, May 2019, Art. no. e11705.



Shaoxiong Ji received the bachelor's degree from the Dalian University of Technology, Dalian, China. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Aalto University, Espoo, Finland.

He was a Research Assistant or a Visiting Researcher with the University of Technology Sydney, Ultimo, NSW, Australia, The University of Queensland, Brisbane, QLD, Australia, and Nanyang Technological University, Singapore. His research interests include machine learning and data mining.



Shirui Pan (Member, IEEE) received the Ph.D. degree in computer science from the University of Technology Sydney (UTS), Ultimo, NSW, Australia, in 2015.

He is currently a Lecturer with the Faculty of Information Technology, Monash University, Melbourne, VIC, Australia. Since 2010, he has published over 80 research articles in top-tier journals and conferences, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), the IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING (TKDE), the IEEE TRANSACTIONS ON CYBERNETICS (TCYB), *Pattern Recognition*, the International Joint Conference on Artificial Intelligence (IJCAI), the International Conference on Data Engineering (ICDE), and the IEEE International Conference on Data Mining (ICDM). His current research interests include data mining, machine learning, and graph data analytics.



Xue Li (Member, IEEE) received the Ph.D. degree from the Queensland University of Technology, Brisbane, QLD, Australia, in 1997.

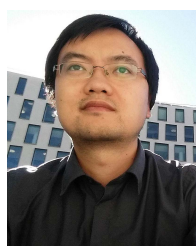
He is currently a Professor with the School of Information Technology and Electrical Engineering (ITEE), The University of Queensland, Brisbane. His current research interests include data mining, social computing, database systems, and intelligent web information systems.



Erik Cambria (Senior Member, IEEE) received the Ph.D. degree from a joint program between the University of Stirling, Stirling, U.K., and the MIT Media Lab, Cambridge, MA, USA.

He worked at Microsoft Research Asia and HP Labs India. He is the Founder of Sentic-Net, Singapore, a Singapore-based company offering B2B sentiment analysis services, and an Associate Professor with Nanyang Technological University (NTU), Singapore, where he also holds the appointment of Provost Chair in Computer Science and Engineering.

Dr. Cambria is involved in many international conferences as a PC member and the program chair. He is also an Associate Editor of several journals, e.g., *Neurocomputing (NEUCOM)*, *Information Fusion (INFFUS)*, *Knowledge-Based Systems (KBS)*, *IEEE Computational Intelligence Magazine (CIM)*, and the IEEE INTELLIGENT SYSTEMS (where he manages the Department of Affective Computing and Sentiment Analysis).



Guodong Long received the Ph.D. degree from the University of Technology Sydney (UTS), Ultimo, NSW, Australia, in 2014.

He is currently a Senior Lecturer with the Australian Artificial Intelligence Institute, Faculty of Engineering and IT, UTS. His research focuses on data mining, machine learning, and natural language processing (NLP). He has more than 40 research articles published on top-tier journals, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), the IEEE

TRANSACTIONS ON CYBERNETICS (TCYB), and the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), and conferences, including the International Conference on Learning Representations (ICLR), the AAAI Conference on Artificial Intelligence (AAAI), the International Joint Conference on Artificial Intelligence (IJCAI), and the IEEE International Conference on Data Mining (ICDM).



Zi Huang (Member, IEEE) received the B.Sc. degree in computer science from Tsinghua University, Beijing, China, and the Ph.D. degree in computer science from the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia.

She is currently 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.