SYSTEMS-LEVEL QUALITY IMPROVEMENT

Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review

Gema Castillo-Sánchez[1] (ID) · Gonçalo Marques[1,2] (ID) · Enrique Dorronzoro[3] (ID) · Octavio Rivera-Romero[3] (ID) ·
Manuel Franco-Martín[4] (ID) · Isabel De la Torre-Díez[1] (ID)

Abstract

According to the World Health Organization (WHO) report in 2016, around 800,000 of individuals have committed suicide. Moreover, suicide is the second cause of unnatural death in people between 15 and 29 years. This paper reviews state of the art on the literature concerning the use of machine learning methods for suicide detection on social networks. Consequently, the objectives, data collection techniques, development process and the validation metrics used for suicide detection on social networks are analyzed. The authors conducted a scoping review using the methodology proposed by Arksey and O'Malley et al. and the PRISMA protocol was adopted to select the relevant studies. This scoping review aims to identify the machine learning techniques used to predict suicide risk based on information posted on social networks. The databases used are PubMed, Science Direct, IEEE Xplore and Web of Science. In total, 50% of the included studies (8/16) report explicitly the use of data mining techniques for feature extraction, feature detection or entity identification. The most commonly reported method was the Linguistic Inquiry and Word Count (4/8, 50%), followed by Latent Dirichlet Analysis, Latent Semantic Analysis, and Word2vec (2/8, 25%). Non-negative Matrix Factorization and Principal Component Analysis were used only in one of the included studies (12.5%). In total, 3 out of 8 research papers (37.5%) combined more than one of those techniques. Supported Vector Machine was implemented in 10 out of the 16 included studies (62.5%). Finally, 75% of the analyzed studies implement machine learning-based models using Python.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Gema Castillo-Sánchez
gemaanabel.castillo@alumnos.uva.es; gemacs17@gmail.com

Gonçalo Marques
goncalosantosmarques@gmail.com

Enrique Dorronzoro
enriquedz@dte.us.es

Octavio Rivera-Romero
octavio@dte.us.es

Manuel Franco-Martín
mfrancom@saludcastillayleon.es

Isabel De la Torre-Díez
isator@tel.uva.es

[1] Department of Signal Theory and Communications, and Telematics Engineering, Universidad de Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain

[2] Polytechnic of Coimbra, ESTGOH, Rua General Santos Costa, 3400-124 Oliveira do Hospital, Portugal

[3] Electronic Technology Department, Universidad de Sevilla, Sevilla, Spain

[4] University Rio Hortega Hospital, Valladolid, Spain

Acronyms and Abbreviations

GC	Gema Castillo
GM	Gonçalo Marques
ED	Enrique Dorronzoro
OR	Octavio Rivera
ML	Machine Learning
GSN	Generic Social Network
OHC	Online Health Community
NPL	Natural Processing Language
LIWC	Linguistic Inquiry and Word Count
LSA	Latent Semantic Analysis
NMF	Non-negative Matrix Factorization
PCA	Principal Component Analysis
CNN-LSTM	Component neural network long short-term memory
MCMC	Markov chain Monte Carlo
ND	Not Defined
SVM	Supported Vector Machine
LR	Logistic Regression
NN	Neuronal Network
KNN	K-Nearest Neighbor
DT	Decision Tree
NB	Naïve Bayes
RoF	Rotational Forest
RF	Random Forrest
CNN	Convolutional Neural Networks
ET	Extra Trees
RNN	Recurrent Neural Network
LDA	Latent Dirichlet Allocation
Km	K-means
AR	Association Rules
PAM	Partitioning Around Medoids
HCA	Hierarchical Clustering Algorithm
BFA	Binary Firefly Algorithm
GBM	Gradient Boost Machines
LiR	Lineal Regression
LSVC	Linear Supported Vector Classifier
P	Perceptron
SGD	Stochastic Gradient Descent
CART	Classification and Regression Tree
SMO	Sequential Minimal Optimization
Bi-LSTM	Bidirectional Long Short-Term Memory
RBF	Radial Basis Function
CHAID	Chi-square Automatic Interaction Detection
MN	Multiple Nominal
MLP	Multilayer Percepton
LM-BFGS	Limited Memory-Broyden–Fletcher–Goldfarb–Shanno
ESGD	Extension of Stochastic Gradient Descent
MCMC	Markov Chain Monte Carlo
SPS	Suicide Probability Scale

Introduction

According to the World Health Organization (WHO) report in 2016, nearly 800,000 people have committed suicide [1]. Suicide is a tragic situation that affects families, neighbours, leaving significant effects on those who survive. It is considered the second cause of unnatural death in people between 15 and 29 years old [2].

The report on "death statistic according to cause of death in Spain" published by the National Statistics Institute, in 2017, the last year for which data is available, states a total of 3679 suicides. Moreover, 140 fewer suicides than the previous year have been reported in 2018 (3539) [3]. The multiple scenarios that families and individuals face in their daily routine can lead to this tragic situation. Consequently, committing suicide is a critical public health challenge that numerous countries address in different manners [4].

Suicidal behaviours are a complex phenomenon that is influenced by multiple factors such as biological, clinical, psychological, and social considerations [5]. On the one hand, suicide is preceded by milder manifestations, such as thoughts of death or suicidal ideation [6]. On the other hand, suicide is closely related to the model of society in which an individual lives [7]. Moreover, it is directly related to the experience of high-stress circumstances and lifestyle changes [8].

Currently, the effects of COVID-19 and isolation will cause a significant emotional impact worldwide [9]. In particular, people who have suffered from mental health diseases are in an even more fragile situation [10]. Therefore, an increase in anxiety and depression disorders, drugs use, loneliness, domestic violence and even suicide are expected to occur in these individuals [11]. Consequently, the risk of suicide attempts has increased among the population [9]. Multiple novel factors contribute to an increase in suicide risk [12]. In particular, the measures for prevention of COVID-19 that includes social distancing plans are strictly related to suicide risk [9].

The reduction in physical contact can lead to loss of protection against suicide [9]. These factors will be even more relevant among people who have previous mental health problems [13]. Social distancing is necessary to control the COVID-19 pandemic and decrease the propagation of the virus [14]. However, a global perspective on indirect mortality is also essential [15]. Social distancing is connected to an increased risk of suicidal behavior [16]. Therefore, social distancing must be addressed through a global intervention plan that implements new models to combat physical distancing using social networks [17]. In this context, several new technologies have been identified as a crucial resource to detect people in suicide risk [18].

Furthermore, young people who constitute a vulnerable group commonly use social networks [19]. Social networks are a popular method of communication between people [20]. Consequently, social networks are an appropriate method to recognize the behaviour of the person according to the content of their posts [21]. The analysis of the user's posts on social media is a complex problem [22]. The complexity is even higher if the objective is to estimate the suicide risk [23]. Also, if the analysis is carried manually by experts, discrepancies usually occur due to the peculiarities of the language used in social networks [24]. Therefore, automatic architectures that use machine learning (ML) methods should be developed.

Nevertheless, numerous of these automated systems require the availability of datasets that allow the training of predictive models which is a critical limitation [25]. On the one hand, these datasets currently do not exist, or they have limited specifications. On the other hand, unsupervised models do not require training. However, these models need datasets for validation [26].

Currently, the use of ML techniques to analyze health-related data is a trending topic. Moreover, the use of different systems based on ML in different areas, such as disease diagnosis and bioinformatics presents promising results [27–33]. In particular, for mental health, various models and tools for suicide risk prevention have been proposed in the literature [34].

This scoping review aims to identify the current ML techniques used to predict suicide risk based on information posted on social networks. This paper reviews the state of the art on this topic focusing on the ML methods, the objectives, the data collection techniques, the development process and the validation metrics used. The main contribution of this study is to summarize the state of the art and to provide a description of the common outcomes and limitations of current research to support future investigations.

The remaining of this paper is organized as follows. Section 2 presents the methodology concerning the search strategy, study selection criteria, screening process, and data extraction. The included studies are analyzed in Section 3 and are discussed in Section 4. Finally, the most relevant findings and the limitations of the study are summarized in Section 5. The PRISMA extension for conducting scoping reviews, the technical details of the machine learning techniques, internal validation strategies and main outcomes of the selected studies are included as supplementary material.

Methodology

Overview

This study summarize the requirements and methods for enhanced suicide risk assessment using social networks.

Consequently, the authors conducted a scoping review using the methodology proposed by Arksey and O'Malley et al. [35]. Furthermore, the authors have followed the PRISMA-ScR proposed by Tricco et al. [36]. The overall procedure is annexed as supplementary material (Appendix I).

On the one hand, Arksey and O'Malley et al. [35] framework is widely used on scoping reviews concerning the health domain. This framework presents relevant recommendations to summarize findings and identify research gaps in the existing literature. On the other hand, the PRISMA extension for scoping reviews built by Tricco et al. [36] defines a checklist of the significant items to be reported when a scoping review is conducted.

Search strategy

The authors have performed a systematic review to identify relevant papers that use suicide risk assessment models in social networks. The search has been conducted during March 11–13 of 2020. The databases used are PubMed, Science Direct, IEEE Xplore and Web of Science since they are the most relevant sources and include the most significant scientific work. The authors have defined the search terms, and the selection of the studies focus on literature written in the English language.

The search string used in the databases was: ["suicide" AND ("social networks" OR "social network") AND "algorithm"].

Study selection criteria

To select the relevant studies on this topic, the authors defined the following inclusion criteria:

- English language.
- Types of studies: Research paper, Clinical Trial and Case Study.
- Published from 2010 up until December 2019.
- Focus on suicide.
- The studies include algorithms or models to estimate suicide risk using the social network.

The research papers were excluded if they were not written in the English language, do not include a specific suicide intervention or do not report information regarding technical aspects of the model/algorithm used to detect suicide risk on social networks.

Screening process

The screening process of the papers obtained through the search strategy was performed by two authors independently (GC and GM). The process was divided into two phases. Firstly, the authors have reviewed the title and abstract. Secondly, the authors have analyzed all the manuscript. The conflicts were resolved by common consensus.

Data extraction

The extraction of the data from the selected studies was performed by four authors (GC, GM, OR and ED). The authors examined the completed form for consistency and accuracy. The extracted data is split into two sets, such as general and technical information. General information refers to the title, year, authors, objectives and methods included in the study. The technical information set is based on Luo et al.'s guidelines [37] and contains the following categories:

- Objectives: refers to the main goals of the proposed ML models. A taxonomy was defined to describe those goals:

  ○ Text Classification: Models that aim to classify post into several categories, including a binary classification, based on post content.
  ○ Entity Recognition: Models that aim to identify several public entities in the text.
  ○ Emotion Recognition: Models that aim to identify emotions expressed in the post content.
  ○ Feature Extraction: Models that aim to collect information regarding characteristics of the post content such as lexical, semantic or sentiment features (word polarity).
  ○ Topics Identification: Models that aim to analyze themes being addressed in the dataset or the posts.
  ○ Features Selection: Models that aim to select automatically features, including optimization and feature reduction, to be included as predictor parameters in the predictive model.
  ○ Score Estimation: Models that aim to estimate a quantitative suicide risk value.

- Data Collection

  ○ Data sources: refers to where the data set for the study is collected. We have followed the taxonomy used by Gonzalez-Hernandez et al. [38]:

  ▪ Generic Social Network (GSN): Social network containing information about a range of topics (e.g. Twitter, Facebook and Instagram).

- Online Health Community (OHC): Domain-specific networks that are dedicated exclusively for discussions associated with health.

  ○ Inclusion and exclusion criteria: information regarding what method was followed to include the data in the data set. The authors define the following possible categories:

  ▪ Keywords: This category includes all studies that defined a set of keywords, hashtags, or phrases to be used as queries or filters.
  ▪ Direct Selection: A set of participants is selected, and then, data from their social networks are included.

  ○ Time spam: refers to the period when data was collected.
  ○ The number of posts: Dataset size referred to as the number of tweets/posts used.
  ○ The number of participants: Number of users/ participants of whom posts were included in the dataset.
  ○ Data description: basic statistics used to describe the dataset in terms of posts classes defined according to the model objectives. For example, the number of suicide risk positive posts, neutral posts, and negative posts.
  ○ Ethics: refers to the inclusion of information about ethical issues and whether ethical approval was obtained from an Ethical Committee for conducting the study.

- Model Development Process

  ○ Data pre-processing: The data preparation techniques such as cleaning, transformation, outliers removal, or missing values processing, were reported in the study.
  ○ Data preparation: The process of Natural Language Processing (NLP) based model development includes a stage in which several parameters are extracted from texts to be used as potential predictor parameters in the model. Most of the techniques used in the data preparation stage are based on data mining. In this category, the authors focus on those data mining techniques for feature extraction, feature detection or entity recognition. Moreover, the authors have collected explicitly mentions about the use of different technologies such as Linguistic Inquiry and Word Count (LIWC), Latent Dirichlet Analysis (LDA), Latent Semantic Analysis (LSA), Non-negative Matrix Factorization (NMF), Word2vec, and Principal Component Analysis (PCA).
  ○ Sentiment analysis: Type of sentiment analysis used in the study.

○ Dataset Annotation: The labelling process followed for dataset annotation. The authors defined the following possible methods:

▪ Manual annotation: The annotation process involved the participation of humans that assessed post contents and assigned one of the possible classes defined.
▪ Corpus: Authors used an existing annotated corpus to train and test the proposed predictive models.
▪ Previous Scores: An assessment using a standard scale or other quantitative instrument was previously conducted. Then, posts were labelled according to the user's score.

○ ML techniques: general ML techniques used in the study.
○ Platform: Platform or language programming used to develop the ML models proposed in the study.

• Internal validation

○ Strategy: How datasets were split into training and testing data.
○ Performance metrics: refers to the metrics used to evaluate the performance of the models
○ Outcomes: refers to the predictive performance of the final model.

Results

The authors retrieved 426 articles in the search conducted in research databases. After removing duplicates, 424 items were selected for screening. The title and abstract review stage resulted in the exclusion of 344 articles since most of the studies do not cumulative focus on suicide risk, social networks and ML methods. After the application of the inclusion and exclusion criteria, 19 papers are included in this work. Three articles were excluded in the full-review stage. One study was excluded since it is based on suicidal behaviour without including a social media analysis [39] . Another study was excluded because it proposes an approach to analyze social media posts for suicide detection, but the authors did not develop any model [40]. Finally, the last exclusion in this stage was conducted since the study proposed by [41] does not include ML techniques.

From the full-text review, 16 articles were then selected for inclusion [26, 42–56]. The flow diagram representing the search process is shown in Fig. 1. Furthermore, the detailed information is presented as supplementary material (Appendix I).

The results of the application of artificial intelligence algorithms or models for suicide risk identification using data collected from social networks have been analyzed in this study.

Furthermore, this paper presents a summary and comparison of the state-of-the-art methods and technical details that address this critical public health challenge.

Description of the included studies

This section introduces a brief description of the articles included in this scoping review.

Ambalavan et al. 2019 [42] developed several methods based on NLP and ML to study the suicidal behaviour of individuals who attempted suicide. The authors built a set of linguistic, lexical, and semantic features that improved the classification of suicidal thoughts, experiences, and suicide methods, obtaining the best performance using a Support Vector Machine (SVM) model.

Birjali et al. 2017 [43] presented a method based on ML classification for the social network Twitter to identify tweets with risk of suicide. The authors used SVM, where SMO (Sequential minimal optimization) is implemented as the best model in terms of precision (89,5%), recall (89,11%) and F-score (89,3%) for suspected tweets with a risk of suicide.

Burnap et al. 2017 [44] developed a set of ML models (using lexical, structural, emotive and psychological features) to classify texts relating to communications around suicide on Twitter. This study presents an improved baseline of the classifier using the Random Forest (RF) algorithm and maximum probability voting classification decision method. Furthermore, the proposed method achieves an F-Score of 72.8% overall and 69% for the suicidal ideation class.

Chiroma et al. 2018 [45] measured the performance of five ML algorithms such as Prism, Decision Tree (DT), Naïve Bayes (NB), RF and SVM, in classifying suicide-related text from Twitter. The results showed that the Prism algorithm had outperformed the other ML algorithms with an F-Score of 84% for the target classes (Suicide and Flippant).

Desmet et al. 2018 [46] have implemented a system for automatic emotion detection based in binary SVM classifiers. The researchers used lexical and semantic features to represent the data, as emotions seemed to be lexicalized consistently. The classification performance varied between emotions, with scores up to 68.86% F-score. Nevertheless, F-scores above 40% was achieved for six of the seven most frequent emotions such as thankfulness, guilt, love, information, hopelessness and instructions.

Du et al. 2018 [47] have investigated several techniques for recognizing suicide-related psychiatric stressors from Twitter using deep learning-based methods and transfer learning strategies. The results show that these techniques offer better results than ML methods. Using a Convolutional neural network (CNN), they have improved the performance of identifying suicide-related tweets with a precision of 78% and an F-1 Score of

Fig. 1 Flow diagram of the scoping review process
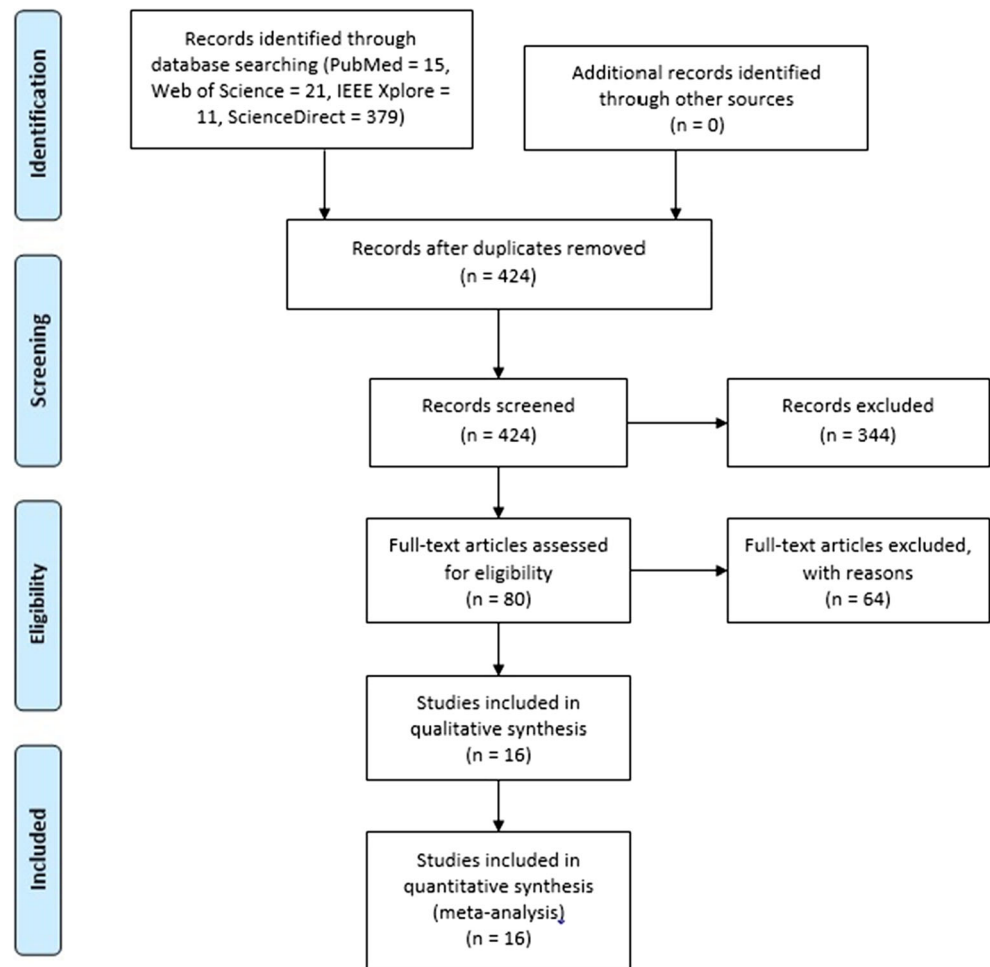
83%, outperforming SVM, Extra Trees (ET), and other ML algorithms. The Recurrent Neural Network (RNN) based psychiatric stressors recognition presented the best F-1 Score of 53.25% by exact match and 67.94% by in-exact match, outperforming Conditional Random Fields (CRF).

Fodeh et al. 2018 [48] proposed a suicidal ideation detection framework that requires a minimum human effort in annotating data by incorporating unsupervised discovery algorithms. This study includes LSA, LDA, and NMF to identify topics. The authors conducted two analysis with k-means clustering and DT algorithms. DT showed better precision (84.4%), sensitivity (91.2%) and specificity (82.9%).

Grant et al. 2018 [49] automatically extracted informal latent recurring topics of suicidal ideation found in social media posts using Word2vec. The proposed method uses descriptive analysis and can identify similar issues to the expert's risk factors.

Jung et al. 2018 [50] have implemented an ontology and terminology method to provide a semantic foundation for analyzing social media data on adolescent depression. They evaluated the ontology obtaining the best values of precision (76.1%) and accuracy (75%) using DT algorithms.

Liu et al. 2019 [51] performed a study to evaluate the feasibility and acceptability of Proactive Suicide Prevention Online (PSPO). PSPO is a new approach based on social media that combines proactive identification of suicide-prone individuals with specialized crisis management. They evaluated different ML models in terms of accuracy, precision, recall and F-measure to get the best performance. The SVM model showed the best performance overall, indicating that PSPO is feasible for identifying populations at risk of suicide and providing effective crisis management.

O'Dea et al. 2015 [52] studied whether the level of concern for a suicide-related post on Twitter could be determined based solely on the content of the post, as judged by human coders and then replicated by ML. They evaluated ML models and decided that the best performing algorithm was the SVM with Term Frequency weighted by Inverse Document Frequency (TFIDF). The results show a prediction accuracy of 76%.

Parraga-Alava et al. 2019 [26] present an approach to categorize potential suicide messages in social media, which is based on unsupervised learning using traditional clustering algorithms. The computational results showed that Hierarchical Clustering Algorithm (HCA) was the best model

for binary clustering achieving average rates of 79% and 87% of F1-score for English and Spanish.

Sawnhey et al. 2019 [53] investigate feature selection using the Firefly algorithm to build an efficient and robust supervised approach for suicide risk detection using tweets. After applying different ML techniques, RF + BFA and CNN-LSTM obtained the best results in accuracy, precision, recall and F1-scores in specific datasets.

Shahreen et al. 2018 [54] used SVM and neural networks (NN) for text classification on Twitter. The researchers used three types of weight optimizers, namely Limited-memory BFGS, Stochastic Gradient Descent and an extension of stochastic gradient descent which is Adam to obtain maximum accuracy. The results show an accuracy of 95.2% using SVM and 97.6% using neural networks. They have used 10-fold cross-validation for model performance evaluation.

Sun et al. 2019 [55] have proposed a hybrid model that combines the convolutional neural network long short-term memory (CNN-LSTM) with a Markov chain Monte Carlo (MCMC) method to identify user's emotions, sample user's emotional transition and detect anomalies according to the transition tensor. The results show that emotions can be well sampled to conform to the user's characteristics, and anomaly can be detected using this model.

Zhang et al. 2014 [56] have used NPL methods and ML models to estimate suicide probability based on linguistic features. The experiments performed by the researchers indicate that the LDA method finds topics that are related to suicide probability and improve the performance of prediction. They obtained the best Root Mean Square Error (RMSE) value of 11 with a linear regression at 1–32 scale.

This paper presents a detailed analysis of the results in the following sections: study objectives, data collection, model development process and this section show data pre-processing, data preparation, sentiment analysis, dataset annotation, ML techniques, platforms and internal validation. The distribution of the included studies according to the year of publication are presented in Fig. 2.
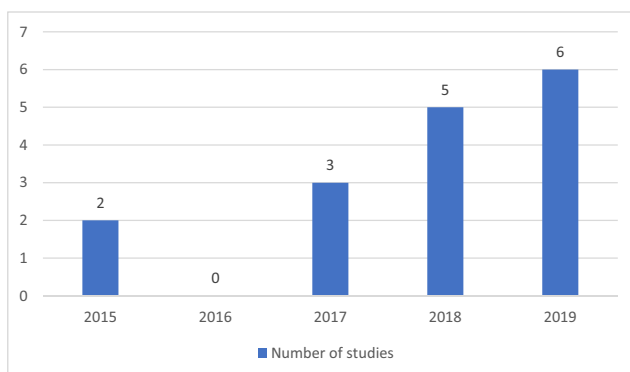

Fig. 2 The distribution of the included studies according to the year of publication

Study objectives

Most of the included studies propose models to classify collected text into suicide-related categories. Text classification is the most common objective in the included studies (12/16, 75%) [26, 42–48, 51–54]. A score estimation of suicidal probability based on post content was proposed in one of the included studies (1/16, 6,25%) [56].

Feature extraction and feature selection were identified as main objectives in four different studies (4/16, 25%) [48, 50, 53, 56]. The remaining categories (Entity Recognition, Theme Identification and Emotion Recognition) were identified only in a study (1/16, 6.25%) [47, 49, 55]. In total, 4 of 16 studies (25%) can be grouped in two categories, involving text classification (3/4) [47, 48, 53] or score estimation (1/4) [56].

Data collection

Different data sources were selected to perform data collection for training and testing of the proposed models. In total, 13 out of the 16 included studies (81.25%) used General social networks (GSNs) for data collection [26, 43–48, 50, 52–56]. The most popular GSN used as the data source in the included studies was Twitter (10/16, 62.5%), followed by forums or microblogs (3/16, 18.75%). Other GSNs used were Weibo (2/16, 12.5%), Facebook, Instagram, Tumblr, and Reddit (1/16, 6.25%). Three studies used OHCs (18.75%), two of them used suicide-related subreddit [42, 49], and the other one used a Sina microblog [51].

Three studies have collected data from OHCs used all posts/comments without defining inclusion/exclusion criteria. Most of the remaining studies defined suicide-related keywords or phrases to filter posts out (10/13, 76.92%) [43–48, 50, 52–54]. Zhang et al. [56], recruited potential participants, and then, the selected participants' posts in Weibo have even used. Finally, two studies that used GSNs did not define inclusion/exclusion criteria (2/13, 15.38%) [26, 55].

The data collection time spam must be reported in ML-based studies as it is defined by the Luo et al.'s guidelines [37]. However, seven of the included studies did not report the time spam when data collection was performed (43.75%) [42, 43, 45, 46, 54–56].

One of the included studies did not report the dataset size (1/16, 6.25%) [54]. The dataset sizes were between 102 posts (minimum) and 1,100,000 posts (maximum). Four out of the remaining 15 studies have used sample sizes between 100 and 999 posts (26.67%) [26, 42–44] . Three of them used sample sizes with more than 800 posts. Five studies reported dataset sizes between 1000 and 5000 posts (33.33%) [45, 47, 50, 52, 53]. Finally, six studies used large dataset, including more than 10,000 posts (40%).

The number of users/participants represented in those datasets was only reported in three studies (18.75%). One of those three

studies recruited 697 participants and then collected data from their Weibo accounts [56]. The other two studies analyzed the user's data collected to report the number of unique users involved in the study ($N = 3873$; $N = 63,252$) [48, 49].

Although using basic statistics to describe dataset is defined as a relevant factor regarding the reliability of ML-based studies in the health domain, as suggested by Luo et al.'s guidelines [37]. However, three of the included studies did not report any dataset description (3/14, 21.43%) [42, 54, 55].

Moreover, only three studies included information regarding ethical issues to collect and manage social media data (3/16, 18.75%). Two of those studies obtained the ethical approval from Ethics Committee: Liu et al. [51] from the Institutional Review Board of the Institute of Psychology, Chinese Academy of Science, and O'Dea et al. [52] from the University of New South Wales Human Research Ethics Committee and the CSIRO Ethics Committee. The remaining study, conducted by Ambalavan et al. [42], adhered to the guidelines defined by Kraut et al. 2004 [57]. It is highlighted that Zhang et al. [56] assessed participants' suicide probability using a standard scale and have collected personal data. However, the information regarding ethical approval was not reported in the article.

Table 1 presents the summary of the results found in terms of the objectives of the study, data sources, ethical aspects, inclusion and exclusion criteria, time span, number of posts, part number and the description of the data of the papers included in this work.

Model development process

Data pre-processing

Data pre-processing is a typical stage in the development process of ML-based models. This stage includes several techniques such as data cleaning, words removal (stop word and punctuation), data transformation, and addressing challenges of outlier or missing values. The reported information regarding data pre-processing is critical for study reproducibility. Most of the included studies reported information regarding the pre-processing stage (14/16, 87.5%) [26, 42, 44–50, 52–56]. Several of these studies only reported vague information and did not include details on the specific techniques and tools used. However, the inclusion of a (sub) section describing data pre-processing is not mandatory. In total, 4 studies included a section/subsection reporting information regarding pre-processing. The remaining studies reported this information in the text. Moreover, some studies presented this information in a different part of the article.

Data preparation

The data mining techniques for feature extraction, feature detection, or entity recognition used in the included studies are summarized in Table 2. In total, 50% of the included studies (8/16) report the use of data mining techniques for feature extraction, feature detection or entity identification [26, 44, 46, 48, 49, 51, 53, 56]. The most common reported technique was LIWC (4/8, 50%), followed by LDA, LSA, and Word2vec (2/8, 25%). Moreover, NMF and PCA were used only in one of the included studies (12.5%). In total, 3 out of 8 studies (37.5%) combined more than one of those techniques.

Sentiment analysis

Seven out of the 16 included studies include sentiment analysis (43.75%). A sentiment ratio or polarity value was assigned to words or features in these studies. Two of these studies used SentiWordNet to obtain the sentiment value [43, 44]. Also, two studies used the categories defined in LIWC as a basis of sentiment value estimation [44, 56]. Furthermore, two studies used previously published lexicons to calculate it [46, 53]. Finally, two studies calculated those values [50, 55], automatically.

Dataset annotation

Supervised learning techniques require labelled, coded, or annotated datasets to train and test the models. In total, 15 out of the 16 included studies required annotated datasets. One of those studies did not report how annotations were performed (6.67%) [54]. Most of the studies followed a manual process to annotate the training and test datasets, involving experts in the codification process (10/15, 66.67%) [42–47, 50–53]. Some of these studies reported detailly how the annotation process was performed. Two studies used existing annotated corpus (13.33%) [26, 55]. In one study (6.67%), the authors designed an algorithm to generate the labels automatically [48]. Finally, a study recruited participants and assessed the participant's suicide probability using a standard scale, the Suicide Probability Scale, and the model results were compared to those obtained using the scale (6.67%) [56].

Machine learning techniques

In total, 15 different ML techniques were used to implement the models proposed in the included studies: Supported Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB), K-means (Km), Deep Learning techniques (DL), Neuronal Network (NN), Linear Regression (LiR), K-nearest neighbour (KNN), Gradient Boost Machine (GBM), Rotational Forest (RoF), Partitioning A M (PAM), Hierarchical Component

Table 1 Data sources and ethics of the included studies

References	Objectives	Data Sources	Ethics	Inclusion / Exclusion criteria	Time spam	Num. posts	Num. part.	Data descrip.
[42]	Text Classification	OHC (subreddit)	Yes	N/A	ND	874	ND	ND
[43]	Text Classification	GSN (Twitter)	No	Keywords	ND	892	ND	Based on the classifier outputs
[44]	Text Classification	GSN (Twitter)	No	Keywords	1 February 2014 to 15 March 2014	816	ND	Class 1: 13% Class 2: 5% Class 3: 30% Class 4: 6% Class 5: 5% Class 6: 15% Class 7: 26%
[45]	Text Classification	GSN (Twitter)	No	Keywords	ND	1060	ND	datasets Binary: Suicide - 156 Flippant - 133 Three classes: Suicide - 156 Flippant -133 Non-Suicide- 771 Seven classes: Suicide - 156 Campaign - 158 Flippant - 133 Support - 178 Memorial - 142 Reports - 165 Other - 128
[46]	Text Classification	GSN (Forums and Netlogs)	No	Keywords	ND	10,040	ND	Training: (N=1000) 851 (82%) relevant; 257 severe 189 irrelevant
[47]	Text Classification+ Entity Recognition	GSN (Twitter)	No	Keywords	26 June 2017 to 19 October 2017	3263 (Classification) 3000 (Recognition)	ND	Classification: 50% in training dataset; The same distribution of original data collected in the evaluation dataset
[48]	Feature Extraction + Text Classification	GSN (Twitter)	No	Keywords	1 January 2015 to 8 June 2016	12,066	3873	dataset 1: 280 users HighRisk 1614 users at risk dataset 2: 280 HighRisk

Table 1 (continued)

References	Objectives	Data Sources	Ethics	Inclusion / Exclusion criteria	Time spam	Num. posts	Num. part.	Data descrip.
								280 AtRisk (randomly selected) dataset 3: 280 HighRisk 285 AtRisk
[49]	Theme Identification	OHC (r/SuicideWatch)	No	N/A	2008–2016	131,728	63,252	N/A
[50]	Features Selection	GSN (Twitter and websites)	No	Keywords	1 January 2012 to 31 December2014	1241	ND	Positive: 506 Negative: 735
[51]	Text Classification	OHC[a] (Zoufan's Sina microblog)	Yes	N/A	Training: 1 to 28 April 2017 Testing: 3 July, 2017 to 3 July 2018	Training: 27,007 Testing: 387,823	ND	Training: Positive: 2786 (10%)
[52]	Text Classification	GSN (Twitter)	Yes	Keywords	18 February 2014 to 23 April 2014	1820	ND	Set A: 27% safe to ignore; 55% possibly concerning; 18% Strongly concerning Set B: 31% secure in ignoring; 58% probably relating to; 11% Strongly concerning
[26]	Text Classification	GSN (Twitter, Facebook, Instagram, and forums)	No	ND	N/A	102	ND	No risk: 70 Urgent: 19 Possible: 8 Immediate: 5
[53]	Features Selection + Text Classification	GSN (Twitter, Tumblr, Reddit) + Forums	No	Keywords	13 January 2018 to 26 March 2018	4314	ND	SCO = 2726 pos + 18,290 neg (training) UNI = 1576 pos + 18,290 neg (training) H = 666 pos + 4130 neg (test)
[54]	Text Classification	GSN (Twitter)	No	Keywords	ND	ND	ND	ND
[55]	Emotion Recognition	GSN (Weibo)	No	N/A	ND	1,100,000	N/A	ND
[56]	Features Extraction + Score Estimation	GSN (Weibo Sina)	No	Direct selection	ND	2000 per user	697	N/A

ND Not defined, *N/A* Not applicable, Data Sources: *GSN* Generic Social Network, *OHC* Online Health Community

[a] A microblog focused on suicide

Table 2 Techniques for feature extraction or selection used in the included studies

Techniques References	LIWC	LDA	LSA	NMF	Word2vec	PCA	Total
[44]	•					•	2
[46]			•				1
[48]		•	•	•			3
[49]					•		1
[51]	•						1
[26]					•		1
[53]	•						1
[56]	•	•					2
Total Number of studies	4	2	2	1	2	1	

LWIC Linguistic Inquiry and Word Count, *LDA* Latent Dirichlet Allocation, *LSA* Latent Semantic Analysis, *NMF* Non-negative Matrix Factorization, *PCA* Principal Component Analysis

Analysis (HCA), and Association Rules (AR). Table 3 shows the distribution of these techniques in the included studies.

SVM was the most used technique being implemented in 10 out of the 16 included studies (62.5%) [42–47, 51–54]. The second most used technique was DT (7/16, 43.75%) [43–45, 47, 48, 50, 51], followed by LR (5/16, 31.25%) [42, 50–53] and RF (4/16, 25%) [45, 47, 51, 53]. DL, NB and Km were used in 3 out of the 16 included models (18.75%). In total, 2 models based on NN were proposed (12.5%) [42, 50]. Finally,

7 of those 15 techniques were used only in a study (LiR, KNN, GBM, RoF, PAM, HCA, and AR).

In total, 25% of the included articles used only a technique to implement the proposed model [46, 49, 55, 56]. The remaining studies developed the proposed models using 2 different techniques (3/16, 18.75%) [48, 52, 54], 3 techniques (3/16, 18.75%) [26, 42, 50], 4 techniques (5/16, 31.25%) [43–45, 47, 51], or 5 techniques (1/16, 6.25%) [53].

Table 3 Machine learning techniques used in the included studies

Techniques	SVM	LR	LiR	DT	NB	KNN	RF	GBM	RoF	Km	PAM	HCA	AR	NN	DL	Total
References																
[42]	•	•												•		3
[43]	•			•	•	•										4
[44]	•			•	•				•							4
[45]	•			•	•		•									4
[46]	•															1
[47]	•			•			•								•	4
[48]				•						•						2
[49]										•						1
[50]		•		•									•			3
[51]	•	•		•			•									4
[52]	•	•														2
[26]										•	•	•				3
[53]	•	•					•	•							•	5
[54]	•													•		2
[55]															•	1
[56]			•													1
Total Number of studies	10	5	1	7	3	1	4	1	1	3	1	1	1	2	3	

SVM Supported Vector Machine, *LR* Logistic Regression, *LiR* Lineal Regression, *DT* Decision Tree, *NB* Naïve Bayes, *KNN* K-Nearest Neighbor, *RF* Random Forrest, *GBM* Gradient Boost Machines, *RoF* Rotation Forest, *Km* K-means, *PAM* Partitioning Around Medoids, *HCA* Hierarchical Clustering Algorithm, *AR* Association Rules, *NN* artificial Neural Network, *DL* Deep Learning

Platforms and internal validation

The platform or software tool used to implement the ML-based models is identified by half of the included studies. Python was the most used tool (6/8, 75%) [26, 42, 46, 49, 52, 54]. One of these studies combines Python and R [26]. Two out of the 8 studies used Weka software to develop the proposed models [43, 44].

One of the included studies focus on topic identification, and authors followed a manual analysis of topics proposed using the models to estimate their validation [49]. Five of the remaining included studies did not report information regarding internal validation strategy followed to assess the validity of the proposed models (33.33%) [26, 43, 48, 50, 55]. The 10-fold cross-validation was the most implemented strategy in the included studies (8/10, 80%) [44–46, 51–54, 56]. One study followed a 70–30 proportion rule to split the dataset in training and test datasets (10%) [42]. However, the technique used to split the data is not reported. Other study followed a 7–1-2 proportion to split the dataset for classifier model validation and a manual selection for classifier validation (10%) [47].

All studies reported the performance parameters used in the validation process. Precision, recall and F-score are the most used performance parameters (12/15, 80%). In total, 66,67% (10/15) of the studies have used accuracy as a performance value. Fodeh et al. [48] used specificity, sensitivity and area under the Receiver Operating Characteristic (ROC) curve (6.67%). Zhang et al. [56] used RMSE value to validate their estimation model.

Discussion

Social networks are an effective method to detect some behaviours. Moreover, they are particularly relevant to identify subjects at suicide risk. The extensive use of social networks leads the authors to investigate the current scenario concerning suicide prevention. This is the primary motivation of the presented research. This study verifies the trends and results of applying ML algorithms and the methods used by various researchers to address this critical situation. Indeed, considering the COVID-19 pandemic, social networks are one of the most used methods of communication. Therefore, it is relevant to survey the main techniques, algorithms and models applied to social networks to detect suicidal risk behaviours.

In total, 43,75% (7/16) of the studies does not provide the time spam information concerning the experiments conducted. This is a relevant limitation, as proposed by Luo et al.'s guidelines [37]. Moreover, 81,25% (13/16) does not specify the number of participants involved. The anonymization of the participant information should be justified. However, it is possible to characterize the participants involved in the

studies and maintain their privacy at the same time. This information allows us to conclude that the quality of the reports of suicide risk prediction models must be increased. The authors must report relevant items to ensure reliability.

Furthermore, the details of the datasets used are not presented in 18,75% (3/16) of the analyzed literature. Although the use of basic statistics to describe dataset is defined as a relevant factor regarding the reliability of ML-based studies in the health domain as proposed by Luo et al.'s guidelines [37]. The dataset description is of utmost importance since the efficiency of the specified results, and their future improvements are closely connected with the sample size.

Furthermore, three studies did not report any dataset description (3/14, 21.43%) [42, 54, 55]. Consequently, it is critical to question what reasons can justify the inexistence of the dataset description. Indeed, this can be related to confidential concerns. However, it is essential to mention that without the complete dataset information is not possible to ensure the absence of bias or deficiencies in the information used. Moreover, it is not possible to ensure the reproduction of the experiments.

In total, 76.92% (10/13) of the studies defined suicide-related keywords or phrases for text analysis. Furthermore, text classification is the objective of 75% of the analyzed studies. Consequently, this denotes a significant limitation concerning the multiple forms of visual communication items such as emoticons that are currently used. However, the reason why most of the authors does not consider the visual components in sentences is not clear. This can be related to technical limitations of the used software tools.

Consequently, it is necessary to promote new research activities to solve this critical limitation. The pre-processing data stage is required to develop or replicate the ML-based model. Therefore, most of the included studies indicated information about the pre-processing stage (14/16, 87.5%) [26, 42, 44–50, 52–56]. Moreover, it should be noted the majority of the studies only present vague information regarding pre-processing data methods and validation strategy. Pre-processing is an essential aspect of detecting suicide risk using ML. However, according to the results achieved, there is a significant limitation related to the unstandardized information of each analyzed research.

Additionally, the authors note that most of the reviewed papers do not present the data processing methods in detail. Consequently, there is a significant limitation concerning the real reason for this scenario. This can be related to methodological or practical difficulties. However, the question about what motivates this trend still exists. Furthermore, there is no justification for this scenario in the before-mentioned studies. Therefore, future research on the subject should ensure the detailed information about the pre-processing methods.

A specific annotated dataset for suicide risk on social media is also a critical limitation. In total, 10 of the 15 papers (75%)

have performed manual annotation. However, it should be noted that the peculiarities of the multiple languages used in social networks can be a relevant limitation for data labelling [38, 58]. The sentiment analysis has been performed in most cases assigning the polarity to the words [59]. However, these polarities could vary according to specific domains such as suicide and considering the terminology used in social networks. Therefore, it is relevant to perform sentiment analysis that encompass the linguistic entities as phrases [60].

Stakeholders have reported several ethical issues as critical factors in the use of social media as a participatory health tool [61]. In this sense, those relevant issues must also be addressed appropriately in ML research applied to the health domain. Despite this relevance, ethics is not appropriately discussed by authors in their reports. There is a lack of information regarding ethical issues in the included studies. Only three studies included information regarding ethical issues to collect and treat social media data (3/16, 18.75%). However, the doubt regarding the reasons that justify the inexistent ethical agreements of the majority of the works still exist. Consequently, a critical limitation is found regarding the ethical concerns involved in the collection and analysis of this sensitive type of data.

Two of those studies obtained an ethical approval from the Ethics Committee ([42, 52]). However, ethical and privacy concerns associated with the data gathering method are a controversial practice. To justify its use, formal prospective studies analyzing if and how physician access to a patient's social media influences care should be performed [62].

Conclusion

This paper has presented a scoping review on the main techniques, algorithms and models applied to social networks to detect suicidal risk. In total, 75% of the included studies propose models to classify collected text into suicide-related categories. Text classification is the main objective of 75% of the included studies. Furthermore, 50% of the included studies (8/16) report explicitly the use of data mining techniques for feature extraction, feature detection or entity identification. The most commonly reported method was LIWC (4/8, 50%), followed by LDA, LSA, and Word2vec (2/8, 25%). NMF and PCA were used only in one of the included studies (12.5%). In total, 3 out of 8 research papers (37.5%) combined more than one of those techniques.

One the one hand, SVM was the most used technique being implemented in 10 out of the 16 included studies (62.5%). On the other hand, the second most used technique was DT (7/16, 43.75%), followed by LR (5/16, 31.25%) and RF (4/16, 25%). The most used platform to implement the ML-based models is Python (6/8, 75%).

Furthermore, all studies reported the performance parameters used in the validation process. Precision, recall and F-score were the most used performance parameters (12/15, 80%). In total, 10 out of 15 studies used accuracy as a performance evaluation metric (66.67%). In summary, ML methods for suicide risk detection and prevention are adjusted to each region, supporting the current pandemic scenario towards enhanced public health and well-being.

Nevertheless, this scoping review has some limitations related to its primary objective. This paper only reviews studies that focus on suicide risks. The papers have been selected using a scoping review methodology in four research databases and written in English. However, other research studies can be available in different languages and databases. Moreover, the authors are aware that are multiple algorithms available bases on statistical assessment. Still, this review only surveys articles that include ML methods to detect suicide risk on social networks.

As future work, several activities can be conducted, such as creating an annotated Corpus for various languages, developing new ML models, especially in other languages than English. These activities aim to classify posts, estimate suicide risk, analyze potential predictive parameters, optimize predictive parameters, and analyze topics considering the temporal component of user posts and specific tools to analyze sentiment.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

Ethical approval   This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. World Health Organization: WHO | Suicide data, http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/, last accessed 2020/07/22.

2. Franco-Martín, M.A., Muñoz-Sánchez, J.L., Sainz-de-Abajo, B., Castillo-Sánchez, G., Hamrioui, S., and de la Torre-Díez, I., A systematic literature review of technologies for suicidal behavior prevention. *J. Med. Syst.* 2018. https://doi.org/10.1007/s10916-018-0926-5.

3. Instituto Nacional de Estadistica: España en cifras 2017, 2017.

4.   Turecki, G., Brent, D.A., Gunnell, D., O'Connor, R.C., Oquendo, M.A., Pirkis, J., and Stanley, B.H., Suicide and suicide risk. *Nat. Rev. Dis. Primers.* 5, 1–22 2019.

5.   Cheng, A.T.A., Chen, T.H.H., Chen, C.C., and Jenkins, R.: Psychosocial and psychiatric risk factors for suicide: Case-control psychological autopsy study. *Br. J. Psychiatry* 177, 360–365 (2000). https://doi.org/10.1192/bjp.177.4.360.

6.   Neeleman, J., De Graaf, R., and Vollebergh, W., The suicidal process; prospective comparison between early and later stages. *J. Affect. Disord.* 82, 43–52 2004. https://doi.org/10.1016/j.jad.2003.09.005.

7.   Neeleman, J.: Beyond risk theory: Suicidal behavior in its social and epidemiological context. *Crisis.* 23, 114–20 2002. https://doi.org/10.1027//0227-5910.23.3.114.

8.   Roy, A., Gene–Environment Interaction and Suicidal Behavior. 2012.

9.   Reger, M.A., Stanley, I.H., and Joiner, T.E., Suicide mortality and coronavirus disease 2019—a perfect storm? *JAMA Psychiatry.* https://doi.org/10.1001/jamapsychiatry.2020.1060. 2020.

10.   Liu, Y., Cao, L., Li, X., Jia, Y., and Xia, H.: Awareness of mental health problems in patients with coronavirus disease 19 (COVID-19): a lesson from an adult man attempting suicide. *Asian J. Psychiatr.* 2020.

11.   Galea, S., Merchant, R.M., and Lurie, N.: The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention, 2020. https://doi.org/10.1001/jamainternmed.2020.1562.

12.   Mannix, R., Lee, L.K., and Fleegler, E.W., Coronavirus Disease 2019 (COVID-19) and Firearms in the United States: Will an epidemic of suicide follow? *Ann. Intern. Med.* 2020. https://doi.org/10.7326/m20-1678.

13.   Smith, K., Ostinelli, E., and Cipriani, A., Covid-19 and mental health: A transformational opportunity to apply an evidence-based approach to clinical practice and research, 2020. https://doi.org/10.1136/ebmental-2020-300155.

14.   Terpos, E., Engelhardt, M., Cook, G., Gay, F., Mateos, M.-V., Ntanasis-Stathopoulos, I., van de Donk, N.W.C.J., Avet-Loiseau, H., Hajek, R., Vangsted, A.J., Ludwig, H., Zweegman, S., Moreau, P., Einsele, H., Boccadoro, M., San Miguel, J., Dimopoulos, M.A., and Sonneveld, P.: Management of patients with multiple myeloma in the era of COVID-19 pandemic: a consensus paper from the European Myeloma Network (EMN). *Leukemia.* 2020. https://doi.org/10.1038/s41375-020-0876-z.

15.   Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M., Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* 2020. https://doi.org/10.1038/s41591-020-0883-7.

16.   West, R., Michie, S., Rubin, G.J., and Amlôt, R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat. Hum. Behav.* 4, 451–459 2020. https://doi.org/10.1038/s41562-020-0887-9.

17.   Lee, S.M., Kang, W.S., Cho, A.R., Kim, T., and Park, J.K., Psychological impact of the 2015 MERS outbreak on hospital workers and quarantined hemodialysis patients. *Compr. Psychiatry* 87, 123–127 2018. https://doi.org/10.1016/j.comppsych.2018.10.003.

18.   Muñoz-Sánchez, J.-L., Delgado, C., Sánchez-Prada, A., Pérez-López, M., and Franco-Martín, M.A., Use of New Technologies in the Prevention of Suicide in Europe: An Exploratory Study. *JMIR Mental Health.* 4, e23 2017. https://doi.org/10.2196/mental.7716.

19.   Zheng, L., Wang, O., Hao, S., Ye, C., Liu, M., Xia, M., Sabo, A.N., Markovic, L., Stearns, F., Kanov, L., Sylvester, K.G., Widen, E., McElhinney, D.B., Zhang, W., Liao, J., and Ling, X.B., Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records.

20.   Muller, E., and Peres, R., The effect of social networks structure on innovation performance: A review and directions for research. *Int. J. Res. Mark.* 36, 3–19 2019.

21.   Kim, J., and Hastak, M., Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manag.* 38, 86–96 2018.

22.   DelPozo-Banos, M., John, A., Petkov, N., Berridge, D.M., Southern, K., LLoyd, K., Jones, C., Spencer, S., and Travieso, C.M. Using neural networks with routine health records to identify suicide risk: Feasibility study. *JMIR Mental Health.* 5, e10144 2018.

23.   Frey, L.M., Hans, J.D., and Cerel, J., Perceptions of suicide stigma: How do social networks and treatment providers compare? *Crisis: The Journal of Crisis Intervention and Suicide Prevention.* 37, 95 2016.

24.   Torous, J., Larsen, M.E., Depp, C., Cosco, T.D., Barnett, I., Nock, M.K., and Firth, J., Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. *Curr. Psychiatry Rep.* 20, 51 2018.

25.   Sawhney, R., Manchanda, P., Mathur, P., Shah, R., and Singh, R., Exploring and learning suicidal ideation connotations on social media with deep learning. Presented at the Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis 2018.

26.   Parraga-Alava, J., Caicedo, R.A., Gomez, J.M., and Inostroza-Ponta, M., An Unsupervised Learning Approach for Automatically to Categorize Potential Suicide Messages in Social Media. Proceedings - International Conference of the Chilean Computer Science Society, SCCC. 2019. https://doi.org/10.1109/SCCC49216.2019.8966443.

27.   Olson, R.S., La Cava, W., Mustahsan, Z., Varik, A., and Moore, J.H., Data-driven advice for applying machine learning to bioinformatics problems. arXiv preprint arXiv:1708.05070. 2017.

28.   Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M., and Notar, M., An application of machine learning to haematological diagnosis. *Sci. Rep.* 8, 1–12 2018.

29.   Sharma, M., Sharma, S., and Singh, G., Performance analysis of statistical and supervised learning techniques in stock data mining. *Data.* 3, 54 2018.

30.   Ngiam, K.Y., and Khor, I.W., Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* 20, e262–e273 2019. https://doi.org/10.1016/S1470-2045(19)30149-4.

31.   Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., and Fotiadis, D.I., Machine learning applications in cancer prognosis and prediction. *Comput. Struc. Biotechnol. J.* 13, 8–17 2015.

32.   Gautam, R., and Sharma, M., Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. *J. Med. Syst.* 44, 49 2020.

33.   Kaur, P., and Sharma, M.: Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: A meta-analysis. *J. Med. Syst.* 43, 204 2019.

34.   Scatà, M., Di Stefano, A., La Corte, A., and Liò, P., Quantifying the propagation of distress and mental disorders in social networks. *Sci. Rep.* 8, 1–12 2018.

35.   Arksey, H., and O'Malley, L., Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol. Theory Pract.* 8, 19–32 2005. https://doi.org/10.1080/1364557032000119616.

36.   Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Lewin, S., Godfrey, C.M., Macdonald, M.T., Langlois, E.V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., and Straus, S.E.,

Transl. Psychiatry 10, 72 2020. https://doi.org/10.1038/s41398-020-0684-2.

PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.*. 169, 467 2018. https://doi.org/10.7326/M18-0850.

37. Luo, W., Phung, D., Tran, T., Gupta, S., Rana, S., Karmakar, C., Shilton, A., Yearwood, J., Dimitrova, N., Ho, T.B., Venkatesh, S., and Berk, M., Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J. Med. Internet Res*. 18, 1–10 2016. https://doi.org/10.2196/jmir.5870.

38. Gonzalez-Hernandez, G., Sarker, A., O'Connor, K., and Savova, G., Capturing the patient's perspective: A review of advances in natural language processing of health-related text. *Yearbook of Medical Informatics*. 26, 214–227 2017. https://doi.org/10.15265/IY-2017-029.

39. Desmet, B., and Hoste, V., Emotion detection in suicide notes. *Expert Syst. Appl.* 40, 6351–6358 2013. https://doi.org/10.1016/j.eswa.2013.05.050.

40. Jindal, S., and Sharma, K., Intend to analyze Social Media feeds to detect behavioral trends of individuals to proactively act against Social Threats. *Proc. Comput. Sci.* 132, 218–225 2018. https://doi.org/10.1016/j.procs.2018.05.191.

41. Cavazos-Rehg, P.A., Krauss, M.J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., and Bierut, L.J., A content analysis of depression-related tweets. *Comput. Hum. Behav.* 54, 351–357 2016. https://doi.org/10.1016/j.chb.2015.08.023.

42. Ambalavan, A.K., Moulahi, B., Azé, J., and Bringay, S., Unveiling online suicide behavior: What can we learn about mental health from suicide survivors of Reddit? *Stud. Health Technol. Inform.* 264, 50–54 2019. https://doi.org/10.3233/SHTI190181.

43. Birjali, M., Beni-Hssane, A., and Erritali, M., Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017). 113, 65–72, 2017. https://doi.org/10.1016/j.procs.2017.08.290.

44. Burnap, P., Colombo, G., Amery, R., Hodorog, A., and Scourfield, J., Multi-class machine classification of suicide-related communication on Twitter. *Online Social Networks and Media.* 2, 32–44 2017. https://doi.org/10.1016/j.osnem.2017.08.001.

45. Chiroma, F., Liu, H., and Cocea, M.: Suicide related text classification with prism algorithm. Proceeding of the 2018 International Conference on Machine Learning and Cybernetics. 2018.

46. Desmet, B., and Hoste, V.: Online suicide prevention through optimised text classification. *Inf. Sci.* 439–440, 61–78 2018. https://doi.org/10.1016/j.ins.2018.02.014.

47. Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., and Xu, H.: Extracting psychiatric stressors for suicide from social media using deep learning. *BMC Medical Inform. Decis. Mak.* 18, 77–87 2018. https://doi.org/10.1186/s12911-018-0632-8.

48. Fodeh, S., Li, T., Menczynski, K., Burgette, T., Harris, A., Ilita, G., Rao, S., Gemmell, J., and Raicu, D.: Using machine learning algorithms to detect suicide risk factors on twitter. IEEE International Conference on Data Mining Workshops, ICDMW. 941–948, 2019. https://doi.org/10.1109/ICDMW.2019.00137.

49. Grant, R.N., Kucher, D., León, A.M., Gemmell, J.F., Raicu, D.S., and Fodeh, S.J., Automatic extraction of informal topics from online suicidal ideation. BMC Bioinformatics. 19, 57–66 2018. https://doi.org/10.1186/s12859-018-2197-z.

50. Jung, H., Park, H.A., and Song, T.M., Ontology-based approach to social data sentiment analysis: Detection of adolescent depression signals. *J. Med. Internet Res*. 19, 2017. https://doi.org/10.2196/jmir.7452.

51. Liu, X., Liu, X., Sun, J., Yu, N.X., Sun, B., Li, Q., and Zhu, T., Proactive suicide prevention online (PSPO): Machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. *J. Med. Internet Res*. 21, 2019. https://doi.org/10.2196/11705.

52. O'Dea, B., Wan, S., Batterham, P.J., Calear, A.L., Paris, C., and Christensen, H., Detecting suicidality on twitter. *Internet Interv.* 2, 183–188 2015. https://doi.org/10.1016/j.invent.2015.03.005.

53. Sawhney, R., Shah, R.R., Bhatia, V., Lin, C.T., Aggarwal, S., and Prasad, M., Exploring the Impact of Evolutionary Computing based Feature Selection in Suicidal Ideation Detection. IEEE International Conference on Fuzzy Systems. 2019. https://doi.org/10.1109/FUZZ-IEEE.2019.8858989.

54. Shahreen, N., Subhani, M., and Mahfuzur Rahman, M., Suicidal Trend Analysis of Twitter Using Machine Learning and Neural Network. 2018 International Conference on Bangla Speech and Language Processing, ICBSLP 2018. 2018. https://doi.org/10.1109/ICBSLP.2018.8554733.

55. Sun, X., Zhang, C., and Li, L., Dynamic emotion modelling and anomaly detection in conversation based on emotional transition tensor. *Inf. Fusion*. 46, 11–22 2019. https://doi.org/10.1016/j.inffus.2018.04.001.

56. Zhang, L., Huang, X., Liu, T., Li, A., Chen, Z., Zhu, T.: Using Linguistic Features to Estimate Suicide Probability of Chinese Microblog Users. Lecture Notes in Computer Science. 8944, 549–559 (2015). https://doi.org/10.1007/978-3-319-15554-8.

57. Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., and Couper, M., Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. American Psychologist. 59, 105–117, 2004. https://doi.org/10.1037/0003-066X.59.2.105.

58. Denecke, K., Information extraction from medical social media. In: Health Web Science. pp. 61–73. Cham: Springer International Publishing, 2015. https://doi.org/10.1007/978-3-319-20582-3_8.

59. Gohil, S., Vuik, S., and Darzi, A., Sentiment analysis of health care tweets: Review of the methods used. *JMIR Public Health Surveill.* 4, e43 2018. https://doi.org/10.2196/publichealth.5789.

60. Denecke, K., Gabarron, E., Grainger, R., Konstantinidis, S.Th., Lau, A., Rivera-Romero, O., Miron-Shatz, T., and Merolli, M., Artificial intelligence for participatory health: Applications, impact, and future implications: Contribution of the IMIA Participatory Health and Social Media Working Group. *Yearb Med Inform*. 28, 165–173 2019. https://doi.org/10.1055/s-0039-1677902.

61. Rivera-Romero, O., Konstantinidis, S., Denecke, K., Gabarrón, E., Petersen, C., Househ, M., Merolli, M., and Mayer, M.Á., Ethical Considerations for Participatory Health through Social Media: Healthcare Workforce and Policy Maker Perspectives: Contribution of the IMIA Participatory Health and Social Media Working Group. Yearb Med Inform. s-0040-1701981, 2020. https://doi.org/10.1055/s-0040-1701981.

62. Pourmand, A., Roberson, J., Caggiula, A., Monsalve, N., Rahimi, M., and Torres-Llenza, V., Social media and suicide: A review of technology-based epidemiology and risk assessment. *Telemedicine and e-Health*. 25, 880–888 2019. https://doi.org/10.1089/tmj.2018.0203.