Artificial Intelligence In Medicine 132 (2022) 102395

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Suicidal behaviour prediction models using machine learning techniques: A systematic review

Noratikah Nordin [a,*], Zurinahni Zainol [a,*], Mohd Halim Mohd Noor [a,*], Lai Fong Chan [b]

[a] School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia
[b] Department of Psychiatry, Faculty of Medicine, National University of Malaysia (UKM) 56000 Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia

A R T I C L E I N F O

Keywords:
Machine learning
Data mining
Suicidal behaviour
Prediction model

A B S T R A C T

Background: Early detection and prediction of suicidal behaviour are key factors in suicide control. In conjunction with recent advances in the field of artificial intelligence, there is increasing research into how machine learning can assist in the detection, prediction and treatment of suicidal behaviour. Therefore, this study aims to provide a comprehensive review of the literature exploring machine learning techniques in the study of suicidal behaviour prediction.
Methods: A search of four databases was conducted: Web of Science, PubMed, Dimensions, and Scopus for research papers dated between January 2016 and September 2021. The search keywords are 'data mining', 'machine learning' in combination with 'suicidal behaviour', 'suicide', 'suicide attempt', 'suicidal ideation', 'suicide plan' and 'self-harm'. The studies that used machine learning techniques were synthesized according to the countries of the articles, sample description, sample size, classification tasks, number of features used to develop the models, types of machine learning techniques, and evaluation of performance metrics.
Results: Thirty-five empirical articles met the criteria to be included in the current review. We provide a general overview of machine learning techniques, examine the feature categories, describe methodological challenges, and suggest areas for improvement and research directions. Ensemble prediction models have been shown to be more accurate and useful than single prediction models.
Conclusions: Machine learning has great potential for improving estimates of future suicidal behaviour and monitoring changes in risk over time. Further research can address important challenges and potential opportunities that may contribute to significant advances in suicide prediction.

1. Introduction

Suicide is a major public health crisis and a significant cause of death worldwide. It is estimated that nearly one million people have died by suicide, and the number of suicide attempts has recently been estimated to be ten to twenty times higher in 2019 [1]. The trend of suicide cases is increasing due to various factors such as economic problems, history of mental health problems, and stressful life events [2]. Therefore, accurate identification and classification of people at risk of suicide are crucial, and suicide prevention is a priority area in global mental health services.

The general term suicidal behaviour is defined as thoughts and behaviours related to a person intentionally taking his or her own life, including suicidal ideation, suicide plan, self-harm, and non-fatal suicide attempt [3]. The phenomena of suicidal behaviour are highly complex and dynamic, encompassing multiple factors such as psychological, biological, environmental and clinical, as well as marked differences between age groups, genders, and geographic regions [4].

Over the past five decades, studies of suicidal behaviour have used conventional statistical techniques to identify, classify and predict an individual's suicide risk [2]. Typically, these techniques produce a simple algorithm that requires researchers to use a limited number of factors to examine the correlation between factors for a simple classification problem and little predictive power to predict suicidal behaviours [5,6]. Given the dynamics and complexity of suicidal behaviours, conventional statistical techniques have hampered effectiveness in informing clinical decision-making in a significant way [7,8].

Research in clinical psychology and psychiatry has recently begun to use data mining and machine learning techniques to overcome the conventional statistical techniques [9–11]. In simple terms, data mining means finding useful patterns in data, and classification is one of the

* Corresponding authors.
E-mail addresses: noratikahnordin@student.usm.my (N. Nordin), zuri@usm.my (Z. Zainol), halimnoor@usm.my (M.H. Mohd Noor).

https://doi.org/10.1016/j.artmed.2022.102395
Received 15 January 2022; Received in revised form 12 August 2022; Accepted 29 August 2022
Available online 6 September 2022
0933-3657/

primary data mining tasks, which involves discovering a predictive learning function that classifies data into one or more classes. In the context of predicting suicidal behaviour, data mining helps in classifying a person with suicidal behaviour into suicide attempters and non-suicide attempters based on all the input values. Data mining is also referred to as machine learning, where a large amount of data is extracted and processed to construct a simple model with valuable utility [12]. Machine learning is the foundation of artificial intelligence, which focuses on teaching computers to learn without the need to be programmed for specific tasks [9]. Machine learning techniques are helpful in classifying large numbers of patients into general risk categories and in identifying potentially at-risk patients whose suicidality might otherwise have gone undetected [11,13]. The benefits of machine learning techniques have the potential to improve the prediction of suicidal behaviour, thereby improving suicide prevention and intervention efforts [14].

A literature review of machine learning techniques for predicting suicidal behaviour is important and timely given the rapid pace of technological developments. Several literature reviews focusing on machine learning techniques for mental health have emerged in the clinical psychology and medical fields. For example, Alonso et al. [15] systematically examined data mining algorithms and techniques in mental health, including dementia, Alzheimer's disease, depression, and schizophrenia; however, this article did not describe the use of machine learning techniques in the context of suicidal behaviour. Shatte et al. [16] reviewed the literature on machine learning and Big Data applications for mental health research and practice and concluded that there is significant room for machine learning applications to improve the efficiency of clinical and research practices in other psychology and mental health contexts. Thieme et al. [17] recently conducted a systematic review of machine learning in mental health from a computer science and human-computer interaction (HCI) perspective. They conclude that machine learning in mental health is still in its infancy and research on the development of machine learning is essential to improve the practicality, acceptability, and effectiveness of the technology in mental health. However, those systematic reviews focus on the application of machine learning approaches for mental health illnesses and are limited for suicidal behaviour studies.

To date, Burke et al. [14] have systematically examined the use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviours in thirty-five papers. They conclude that predicting and preventing suicide using machine learning techniques can improve predictive accuracy and identify new indicators of suicidal behaviour. Some opportunities and challenges of machine learning in the study of suicidal behaviour were also briefly discussed. However, they were not explored in detail, and different types of machine learning techniques based on model performance were not presented. Previous reviews have shown that machine learning techniques are scalable, robust and effective for mental health applications, but the literature on suicidal behaviour prediction techniques and the challenges associated with them is still quite limited [2,14,18]. Therefore, this review aims to provide a comprehensive overview of the literature with regard to machine learning techniques in the study of suicidal behaviour prediction. This review would help inform practitioners and researchers about the methods and applications of machine learning in the field of clinical psychology, particularly in studies on suicidal behaviour and highlight gaps as well as potential opportunities for further research.

In the Methodology section, we first explain the search strategies we used to find relevant literature. Next, we perform a synthesis of the literature and describe the use of machine learning techniques to predict suicidal behaviour in each article. Finally, we summarize the findings and provide a set of suggestions and recommendations for further research to develop and improve machine learning techniques that are practically useful in assisting and preventing potentially fatal outcomes of suicidal behaviour.

2. Methodology

The systematic review was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) method proposed by [19]. The PRISMA method is the best method to guide authors in conducting systematic reviews and meta-analyses in a structured way as a guide [20]. A systematic review with an accurate and comprehensive investigation can analyze different ideas published in conventional articles by various researchers. The essential part is to determine the eligibility criteria, which should be carefully selected to describe the hypothesis [21]. According to PRISMA guidelines, the following sections include literature search, identification, screening, and eligibility. The PRISMA checklist of 27 items was selected in this paper to improve the quality of the process. This checklist was developed to increase the accuracy of all reviewed articles in this paper. This method is currently known as one of the best standards for reviewers in reporting their findings [22].

2.1. Literature search

In this step, four well-established scientific databases were selected as appropriate databases to be searched to find relevant articles based on the research questions. Given the benefits of implementing machine learning in healthcare, the current study aims to conduct a systematic review of articles that use machine learning techniques to improve the identification and prediction of suicidal behaviour. Thus, the research questions for this study are (1) What types of machine learning techniques are currently being developed for suicidal behaviour prediction models? (2) What features/indicators have been used to develop predictive models of suicidal behaviour using machine learning techniques? and (3) What are the potential opportunities and challenges in developing the predictive models using machine learning techniques?. Searches of primary research for this systematic review were conducted on Web of Science, PubMed, Dimensions, and Scopus. These databases are widely used by computer science, clinical informatics and health services researchers. It was recommended that more than one database be used to increase the likelihood of finding appropriate and relevant articles for this systematic review [23].

2.2. Identification

The first step in the process of systematic review is identification. This step is about determining the right keywords and developing the search string. According to the defined research question and the final objective, the literature search was completed by using the keywords, including (a) “data mining” or “machine learning”; (b) “suicide” or “suicidal behaviour” or “suicide attempt” or “suicide ideation” or “suicide plan” or “death by suicide” or “self-harm”. Suicidal behaviour is defined as thoughts and behaviours related to an individual intentionally taking their own life including suicide death, suicide attempt, suicide ideation, suicide plan and self-harm [24]. Suicide refers to an act in death which is initiated and carried out by an individual to the end of the action (fatal) while a suicide attempt refers to an act in which an individual harms herself or himself with the intention to die and survive (non-fatal) [3]. Suicide ideation refers to an individual having thoughts of suicide with or without taking their own life, and a suicide plan defines as a formulation (thoughts) of how and when to perform a suicidal act without active preparation [23]. Self-harm refers to a non-fatal act in which an individual harms herself or himself intentionally or without intentionally with varying motives including the wish to die [23]. In this systematic review, we use the general term of suicidal behaviour referring to all the possible suicidal thoughts and behaviours. Published studies were then searched and identified using a search strategy developed by the reviewers. The search strategy was written separately for each database and is presented in Table 1. Articles were searched from January 2016 to September 2021. As a result, 230 records were

Table 1
Search strategy in a different database.

Databases	Search strategy	Articles
Web of Sciences	TITLE-ABS-KEY ("data mining" OR "machine learning" AND "suicide*") AND (LIMIT-TO (PUBYEAR, 2016–2021) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))	66
PubMed	Search: ("data mining"[All Fields] OR "machine learning"[All Fields]) AND "suicidal behaviour"[All Fields] OR "suicid*") AND (2016:2021[pdat]) AND ((fft [Filter]) AND (English[Filter]))	45
Dimensions	Search: ("data mining" OR "machine learning") AND ("suicid*") AND (pubtype:article) AND ("2016:2021")	43
Scopus	Search: (("data mining" OR "machine learning") AND ("suicidal behaviour" OR "suicide" OR "suicide attempt") AND (pubtype:article) AND ("2016–2021") AND ("English")	76
Total		230

found based on the search strategy.

2.3. Screening

Screening was performed by establishing inclusion and exclusion criteria to remove duplicate or irrelevant articles (see Table 2). All articles were screened, and out of 230 articles, 123 articles were removed. 107 articles were selected after removing duplicates and irrelevant articles. According to the exclusion criteria, the eligible articles were selected, and the book chapters, systematic reviews, dissertations and theses, short reports and non-English language papers were removed. In addition, we selected several inclusion criteria such as subject area, document type, years, source type, language and access type as search formulations. A total of 67 articles met the inclusion criteria relevant to the research question of this systematic review.

2.4. Eligibility

Finally, in the last steps, we reviewed the full text of 67 articles to obtain the final collection of studies that will contribute to the review. The included articles were fully reviewed, mainly to extract and summarize the required information, with the aim of answering the main research questions. The articles were precisely selected based on the previously established criteria and reviewed one by one. Thirty-two full-text articles were rejected and excluded because this work did not use data mining and machine learning techniques to identify and predict the suicidal behaviour of individuals. The studies used conventional statistical techniques such as chi-square tests, multivariate analysis and Mann-Whitney U analysis. Finally, 35 articles were found ready and relevant for further analysis, as illustrated in Fig. 1. The PRISMA method and the selection of the right articles took a lot of time. However, the specific structuring of the method ensured that the most appropriate and relevant articles to the main topic of the systematic review were selected.

Table 2
Inclusion and exclusion criteria for selecting eligible articles.

Criteria	Inclusion	Exclusion
Years	January 2016–September 2021	Publications before January 2016
Language	English	Non-English
Source type	Journals	Other than journals
Document type	Articles	Other than articles
Access type	Open access/subscription articles	Other than open access/ subscription articles
Subject area	Computer Sciences, Healthcare Sciences, Healthcare Informatics	Other than Computer Sciences, Healthcare Sciences, Healthcare Informatics

3. Results

3.1. Data extraction

Of the 230 studies that the search yielded, 35 articles met the inclusion criteria. The distribution of recent studies from January 2016 to September 2021 related to the use of machine learning techniques in the study of suicidal behaviour prediction is shown in Fig. 2. The 35 articles were further evaluated to find out what has been developed to classify and predict suicidal behaviour using machine learning techniques. Table 3 represents the details of the identified studies, including the countries of the articles, description of the sample, sample size, outcome, study design, classification types, number of features used to develop the models, types of machine learning techniques, and evaluation of the performance metrics.

3.2. Types of machine learning techniques for suicidal behaviour prediction

Our systematic review identified eight types of data mining and machine learning techniques in the study of suicidal behaviour, namely Bayesian-based, instance-based, neural network, regularization, decision tree, support vector machine, regression, and ensemble techniques as shown in Fig. 3.

Bayesian-based classification is one of the machine learning techniques that make probabilistic predictions based on Bayes' theorem. In probability theory, Bayes' theorem describes the probability of an event based on prior knowledge about the conditions that might be associated with the event [58]. In the context of predicting suicidal behaviour, if a higher risk of suicide attempts is related to age and gender. Then, a person's age and gender can be used to estimate the likelihood of suicide attempts using Bayes' theorem more accurately than estimate without knowledge of a person's age and gender. Three studies used a Bayesian-based model to predict suicide attempts, namely Bayesian network and Naïve Bayes. Oh et al. [40] utilized the Bayesian network to predict suicidal ideation in the Korean population and found that the Bayesian network performed better compared to conventional logistic regression. Barak-Corren [28] applied Naïve Bayes for predicting suicidal behaviour in healthcare centre and Nordin et al. [50] also proposed Naïve Bayes to predict suicidal attempts and showed that this model can predict suicidal behaviour with moderate accuracy (accuracy = 0.82).

Instance-based learning, also known as k-nearest neighbours (KNN), is one of the machine learning techniques used for predicting suicidal behaviour. KNN is a non-parametric and non-linear classification in which all available cases are stored and new cases are classified based on a similarity measure (distance measure). Cases are then classified by a majority vote between the nearest K neighbours [27,57]. Based on the analysis, we identified five studies that apply k-nearest neighbours in classifying suicidal behaviour [27,48,50,51,57]. Compared to other machine learning techniques, KNN achieved moderate accuracy in the range of 0.73 to 0.89 in most of the identified studies.

Moreover, a neural network or artificial neural network (ANN) is a machine learning technique modelled loosely after the human brain. It consists of a network of artificial neurons that process data in a connectionist manner. In ANN, a neuron is a basic unit of the network. We found that six studies used neural networks [25,32,36,40,45,46] and one study used a deep neural network [33] to predict suicidal behaviour. Most of the studies reviewed that the neural network achieved a low accuracy of 0.62 to 0.78. The study by Horvath et al. [46] showed that the neural network did not perform as expected because the study included a limited number of samples, and the neural network is more suitable for inputs based on a continuous value rather than categorical values. Choi et al. [33] found that the deep neural network was less accurate (0.683) in predicting suicide compared to Cox regression and support vector machine.

Our systematic review identified twelve studies that used decision

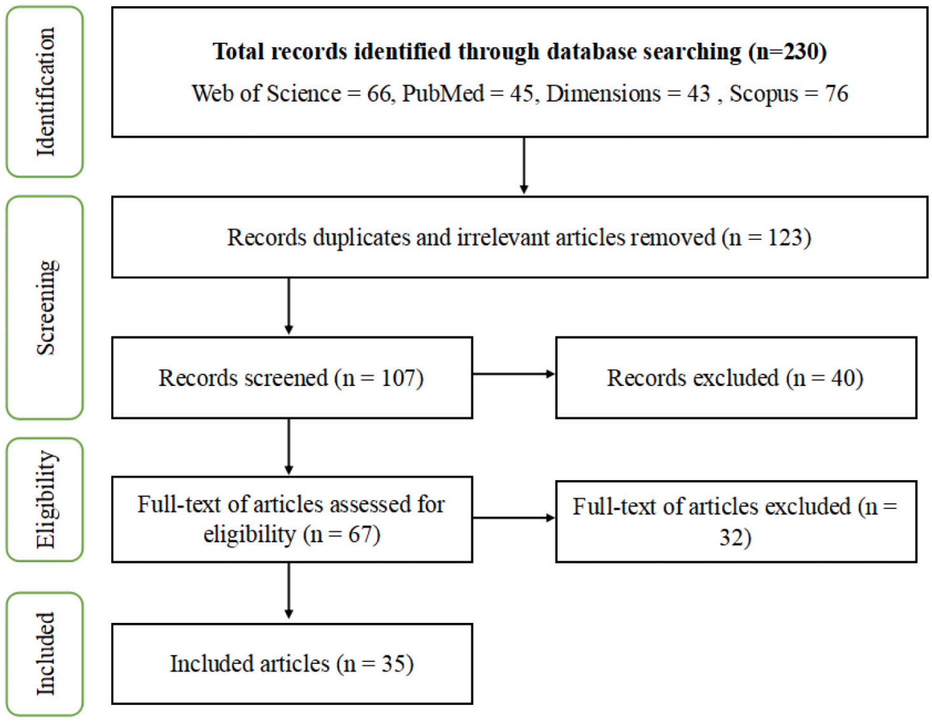

Fig. 1. PRISMA diagram for the identification, screening, eligibility, and included articles.

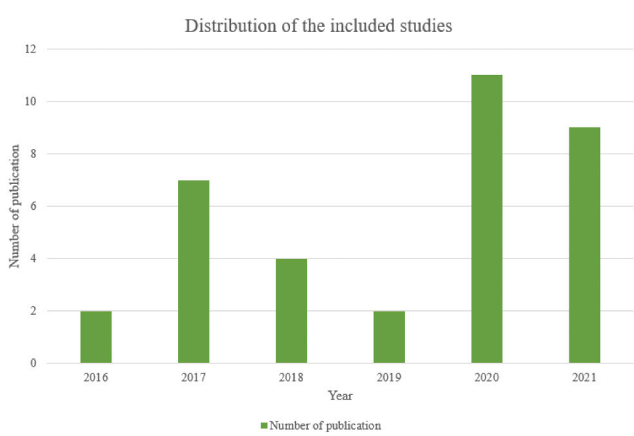

Fig. 2. Distribution of included studies.

trees as machine learning techniques to classify and predict suicidal behaviour [27,40,42–44,46]. A decision tree is a non-parametric and non-linear classification defined as a classification scheme that generates a tree and a set of rules from a given dataset. Decision trees are recursively generated from a dataset to form 'nodes' and 'leaves'. The nodes (also called decision nodes) define splitting conditions for features, and the leaves, called terminal nodes, are labelled with a class. Classification starts from the root node to one of the leaves by moving from one node to another [50,58]. In the context of predicting suicidal behaviour, the leaves represent the suicide risk (yes or no) to be inferred, while the nodes reflect the features that lead to this suicide risk in suicidal behaviour. Edgcomb et al. [49] proposed a classification and regression tree (CART) for predicting suicidal behaviour and self-injury in adults with severe mental illness. The results showed that CART is able to predict the risk of suicide attempt with good performance (accuracy – 0.80, AUC = 0.86, sensitivity = 0.79, specificity = 0.81). In addition, most studies indicated that decision tree classification techniques are able to classify a person with suicidal behaviours with moderate performance in terms of accuracy (0.72–0.91), sensitivity (0.60–0.90) and specificity (0.50–0.85).

Support Vector Machine (SVM) is a machine learning technique to find optimal hyperplanes that separate any class of input spaces. The optimal hyperplanes have a maximum margin between the classes, resulting from discriminant boundaries (dividing lines) [25,27,50]. In the case of predicting suicidal behaviour, there are two classes, namely predicted suicide attempters and predicted non-suicide attempters. By using a variety of kernels and parameters, statistical importance weights can be assigned to the features used. We identified twelve studies that used support vector machine to predict suicide attempters and non-suicide attempters [25–27,30,31,33,36,40,43,48,50,51]. Based on the identified studies, the SVM was able to classify suicide attempters with moderate accuracy (0.78–0.84). Amini et al. [25] found that the SVM performed better than other models (decision tree, Naïve Bayes, logistic regression). In the study by Nordin et al. [50], they also indicate that SVM is the best classifier for predicting suicide attempters in a single predictive model compared to other single predictive models (KNN, logistic regression, decision tree).

Thirteen studies utilized regression when examining suicidal behaviour prediction [25,31,33,36,38,40,41,43,46–48,50,53]. Of the thirteen studies, twelve studies used logistic regression and one study used a Cox regression [33] to develop predictive models for an individual with suicide deaths. Choi et al. [33] used Cox regression to investigate the hazard ratios for suicide and build a predictive model for the likelihood of suicide death within ten years. The results showed that Cox regression has a good AUC performance of 0.688 compared to support vector machine (0.687) and deep neural network (0.683). Logistic regression is the common technique used for machine learning in predicting suicidal behaviour. It is defined as a classification model used to assign observations to a discrete group of classes, for example, suicide attempters or non-suicide attempters [25,50]. Logistic regression transforms its output using a more complex cost function known as a logistic sigmoid function to obtain a probability value [46,47]. Most studies have shown that logistic regression is able to classify and predict suicidal behaviour in an individual with moderate predictive power (AUC = 0.74–0.877, accuracy = 0.64–0.83, specificity = 0.58–0.85,

Table 3
Overview of included studies.

No	Studies	Countries	Sample population	Sample size	Outcome	Study design	Classification types	No of features	Machine learning techniques	Performance matric(s)
1.	Amini et al. (2016) [25]	Iran	General population in Hamadan Province	5414	Suicide deaths	Longitudinal	Binary	6	Logistic regression, Support vector machine, Decision tree, Artificial neural network (CV: training 70 %, 30 % testing)	Sensitivity – 0.72–0.88 Specificity – 0.46–0.67 Accuracy – 0.50–0.68 AUC – 0.719–0.752
2.	Passos et al. (2016) [26]	United States	Patients with mood disorder/ bipolar disorder	144	Suicide attempts	Cross-sectional	Binary	16	Support vector machines, Relevance vector machine, Least absolute shrinkage and selection operator (CV: 10-fold, leave-one-out)	Accuracy – 0.72, Sensitivity – 0.72, Specificity – 0.71
3.	Barros et al. (2017) [27]	Chile	Patients with mood disorders from outpatient and inpatient mental health	707	Suicidal behaviours	Cross-sectional	Binary	343	Decision tree, K-nearest neighbour, Random forest, AdaBoost, Support vector machine (CV:10-fold)	Accuracy – 0.78, Sensitivity – 0.77, Specificity – 0.78
4.	Barak-Corren (2017) [28]	United States	Outpatient and inpatient healthcare centre	20,246	Suicidal behaviours	Longitudinal	Binary	10	Naïve Bayes	AUC – 0.77 Specificity – 0.90 Sensitivity – 0.45 PPV – 0.04
5.	Gradus et al. (2017) [29]	United States	Veterans	2088	Suicide ideations	Cross-sectional	Binary	25	Decision tree, Random forest (CV: bootstrapping)	AUC – 0.92
6.	Hettige et al. (2017) [30]	Canada	Patients with schizophrenia spectrum disorder	345	Suicide attempts	Cross-sectional	Binary	27	Least absolute shrinkage and selection operator, Random forest, Support vector machine, Elastic net (CV: 10-fold stratified)	AUC – 0.67-0.71 Accuracy – 0.65–0.67 (LASSO) Specificity – 0.68, Sensitivity – 0.64
7.	Kessler et al. (2017) [31]	United States	Veterans	6360	Suicide deaths	Longitudinal	Binary	61	Elastic net, Decision tree, Random forest, Adaptive splines, Adaptive boosting, Support vector machine (CV: internal independent test)	Sensitivity – 2.2–26.3
8.	Oh et al. (2017) [32]	South Korea	Patients from outpatient mental health care	573	Suicide attempts	Cross-sectional	Binary	41	Artificial neural network (CV: 70 % training, 15 % validation, 15 % testing)	Accuracy – 0.87 Specificity – 0.91 Sensitivity – 0.78 AUROC – 0.89
9.	Walsh et al. (2017) [5]	United States	Patients at a community hospital	5167	Suicide attempts	Longitudinal	Binary	1328	Random forest (CV: bootstrapping, holdout set)	AUC – 0.84, Precision – 0.79, Recall – 0.95, Brier score – 0.14 Prediction window (7, 14, 30, 60, 90, 180, 365, 720)
10.	Choi et al. (2018) [33]	South Korea	Korean population	819,951	Suicide deaths	Longitudinal	Binary	13	Cox regression, Support vector machine, Deep neural network. (CV: 70 % training, 30 % validation).	AUC Cox regression – 0.688 AUC SVM – 0.576 AUC DNN – 0.632
11.		South Korea	Korean population	11,628	Suicide ideations	Longitudinal	Binary	15	Random forest (CV: 10-fold)	AUC – 0.85 Accuracy –

(continued on next page)

Table 3 (*continued*)

No	Studies	Countries	Sample population	Sample size	Outcome	Study design	Classification types	No of features	Machine learning techniques	Performance matric(s)
	Ryu et al. (2018) [34]									0.821 Sensitivity – 0.836 Specificity – 0.807 Positive predictive value – 0.462 Negative predictive value – 0.961
12.	Simon et al. (2018) [35]	United States	Patients from mental health speciality and primary care	2,960,929	Suicide attempts and suicide deaths	Longitudinal	Binary	313	Logistic regression with least absolute shrinkage and selection operator (CV: 65 % training, 35 % validation)	Sensitivity – 16.8–92.1 Specificity – 50.0–99.1
13.	Walsh et al. (2018) [8]	United States	Adolescents	2247	Suicide attempts	Longitudinal	Binary	100	Random forest (CV: bootstrapping, holdout set)	AUC – 0.83–0.97 Prediction window (7, 14, 30, 60, 90, 180, 365, 720)
14.	Jung et al. (2019) [36]	South Korea	Korean adolescents	59,984	Suicide attempts and suicide ideations	Cross-sectional	Binary	26	Logistic regression, Random forest, Support vector machine, Artificial neural network, Extreme gradient boosting (CV: 5-fold)	AUC – 0.851–0.863 Accuracy – 0.77–0.79 Specificity – 0.77–0.79 Sensitivity – 0.782–0.785
15.	Ribeiro et al. (2019) [37]	United States	Adults from online Web forums	1021	Suicide ideations	Longitudinal	Binary	51	Random forest (CV: bootstrapping) Prediction – 3, 14 28 days	AUC – 0.47–0.84 Precision – 0.76–0.96 Recall – 0.52–0.97
16.	Su (2020) [38]	United States	Children and adolescents from the children medical centre	41,721	Suicidal behaviours	Longitudinal	Binary	30	Logistic regression L1 penalized regularization, logistic regression with sequential forward selection (CV: 90 % training, 10 % testing, 5-fold validation)	AUC – 0.81–0.86 Specificity – 0.90. Predicted window (days – 0, 7, 14, 30, 60, 90, 180, 270, 365)
17.	Shen et al. (2020) [39]	China	Medical college students	4882	Suicide attempts	Cross-sectional	Binary	37	Random forest (CV: 5-fold)	AUC – 0.9255 Accuracy – 0.90 Sensitivity – 0.73 Specificity – 0.91
18.	Oh et al. (2020) [40]	Korea	Korean general population	20,225	Suicide ideations	Cross-sectional	Binary	21	Bayesian network, LogitBoost with logistic regression, Support vector machine, Decision tree, Artificial neural network (CV: 81.3 % training, 18.7 % testing, 5-fold)	AUC – 0.794–0.877 Accuracy – 0.71–0.788 Sensitivity – 0.77–0.819 Specificity – 0.71–0.812
19.	Miche et al. (2020) [41]	Germany	Adolescents and young adults in the community	2797	Suicide attempts	Longitudinal	Binary	16	Logistic regression, Lasso, Ridge, Random forest (CV: 10-fold)	AUC – 0.824–0.829 Sensitivity – 0.028–0.251
20.	Van Mens et al. (2020) [42]	Scotland	Young adults within Scotting wellbeing study	3508	Suicide attempts and suicide ideations	Longitudinal	Binary	14	Logistic regression, K-nearest neighbours, Decision tree, Random forest, Gradient boosting, Support vector machine (CV: 70 % training, 30 % testing, 10-fold)	AUC – 0.74–0.81 Accuracy – 0.72–0.76 Sensitivity – 0.6–0.71 Specificity – 0.74–0.87 Positive PV – 0.29–0.43

(*continued on next page*)

Table 3 (*continued*)

No	Studies	Countries	Sample population	Sample size	Outcome	Study design	Classification types	No of features	Machine learning techniques	Performance matric(s)
21.	Lin et al. (2020) [43]	Taiwan	Military personnel	3546	Suicide ideations	Cross-sectional	Binary	6	Logistic regression, Decision tree, Support vector machine, Random forest, Gradient boosting, Multilayer perceptron (CV: 10-fold)	AUC – 0.88–1.0 Accuracy – 0.94–1.0 Sensitivity – 0.77–1.0 Specificity – 0.80–1.0 Precision – 0.80–1.0
22.	Burke et al. (2020) [44]	United states	Youth at emergency department and primary care	13,325	Suicide attempts	Cross-sectional	Binary	53	Decision tree, Random forest	Accuracy – 0.92–0.99 Sensitivity – 0.52–0.74 Specificity – 0.94–0.975 Precision – 0.43–0.557
23.	Chen et al. (2020) [45]	Sweden	Inpatient and outpatient psychiatric speciality care	541,300	Suicide attempts and suicide deaths	Longitudinal	Binary	425	Elastic net penalized logistic regression, Random forest, Gradient boosting, Neural network (CV: 80 % training, 20 % testing, 10-fold)	AUC – 0.88–0.89 Sensitivity – 0.95–0.96 Specificity – 0.959–0.966
24.	Horvath et al. (2020) [46]	Australia	US prisoners population	353	Suicide attempts	Cross-sectional	Binary	29	Gradient boosting, Neural network, Random forest, Decision tree, Logistic regression, Linear regression (CV: 78 % training, 22 % validation).	AUC – 0.579–0.955 F1 score – 0.417–0.846 Precision – 0.389–0.917 Sensitivity – 0.357–0.786
25.	Iorfino et al. (2020) [47]	Australia	Young people at youth mental health services	1962	Self-harm (suicide attempts and non-suicidal self injury)	Longitudinal	Binary	37	Random forest, Boruta, Lasso regression, Elastic net regression, Bayesian additive regression, Logistic regression (CV: 10-fold)	AUC – 0.744–0.755 Specificity – 0.686–0.722 Sensitivity – 0.684–0.752
26.	van Mens et al. (2020) [48]	Netherlands	Patients at Nivel primary care	207,882	Suicide attempts	Longitudinal	Binary	20	Random forest (CV:70 % training, 30 % testing, 10-fold)	AUC – 0.82 Sensitivity – 0.39 Specificity – 0.98 Accuracy – 0.68
27.	Edgcomb et al. (2021) [49]	United States	Adults patient with depression, bipolar and psychotic disorders	15,644	Suicidal behaviours	Longitudinal	Binary	23	Classification and Regression Tree (CART) (CV: 10-fold)	AUC – 0.86 Sensitivity – 0.79 Specificity – 0.81 Accuracy – 0.80
28.	Nordin et al. (2021) [50]	Malaysia	Patients with depression	75	Suicide attempts	Cross-sectional	Binary	15	Logistic regression, decision tree, Support vector machine, Naïve Bayes, k-nearest neighbours, Random forest, Bagging and Voting. (CV: 3-fold)	AUC – 0.65–0.87 Accuracy – 0.79–0.92 Sensitivity – 0.86–0.92 Specificity – 0.50–0.58
29.	Kirlic et al. (2021) [51]	United States	First-university students	356	Suicidal thoughts and behaviours	Cross-sectional	Binary	49	Elastic net, Support vector regression, Random forest, k-nearest neighbours (CV: 5-fold)	28.3 % of variance (95 % CI: 28–28.5 %)
30.	Macalli et al. (2021) [52]	France	Community sample of college students.	5066	Suicidal thoughts and behaviours	Longitudinal	Binary	70	Random forest (CV: 10-fold)	AUC – 0.80 Sensitivity – 0.79
31.	McMullen et al. (2021) [53]	United States	High-risk inpatients	591	Suicide ideations	Cross-sectional	Binary	59	Random forest, Logistic regression, Gradient boosted trees (CV: NA)	AUC – 0.82–0.90 Accuracy – 0.95-0.98 Precision – 0.57–0.98

(continued on next page)

Table 3 (*continued*)

No	Studies	Countries	Sample population	Sample size	Outcome	Study design	Classification types	No of features	Machine learning techniques	Performance matric(s)
										Recall – 0.16–0.48
32.	Van Vuuren et al. (2021) [54]	Netherlands	General population of students in secondary education	8888	Suicidal behaviours	Longitudinal	Binary	30	Random forest, Lasso regression, Decision rule (70 % training, 30 % testing, CV: 10-fold)	AUC – 0.64–0.79 Sensitivity – 0.34–0.52 Specificity – 0.85–0.94 Accuracy – 0.64–0.68
33.	Cho et al. (2021) [55]	South Korea	Elderly population	48,047	Suicide deaths	Longitudinal	Binary	22	Random forest (CV: 70 % training, 30 % testing, 10-fold)	AUC – 0.818 Accuracy – 0.832 Sensitivity – 0.600 Specificity – 0.833
34.	Navarro et al. (2021) [56]	Canada	Adolescents and young adulthood	1623	Suicide attempts	Longitudinal	Binary	150	Random forest (CV:80 % training, 20 % testing)	Specificity – 0.76 Sensitivity – 0.50 AUC – 0.72
35.	Kim et al. (2021) [57]	South Korea	College students	7824	Suicide attempts and suicide ideations	Cross-sectional	Binary	50	Random forest, K-nearest neighbours (CV: 80 % training, 20 % testing)	AUC – 0.639–0.851 Precision – 0.897–0.953 Recall – 0.916–0.950

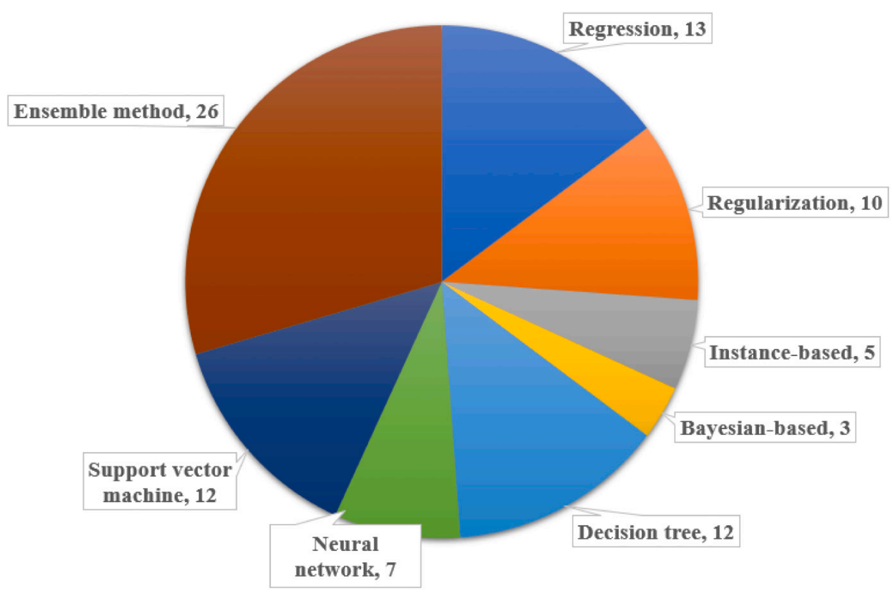

Fig. 3. Machine learning techniques in the study of suicidal behaviour prediction.

sensitivity = 0.655–0.90).

Regularizations are techniques used to minimize the error by properly fitting a function to the given data set and avoid overfitting [58]. We identified several types of regularization techniques, namely L_1 regularization or Least Absolute Shrinkage and Selection Operator (LASSO), L_2 regularization or Ridge regression, and Elastic net regularization. Lasso regression (LASSO) implements a standard logistic regression formula but adds a 'penalty procedure' by assigning some feature weights as zero (adds a penalty term to the cost function), while ridge regression adds a penalty term which is equal to the square of the coefficients [58]. Elastic net regularization is known as a hybrid approach using both the L_1 penalty of Lasso regression and the L_2 penalty of Ridge regression. Based on the analysis, ten studies apply regularization techniques to classify a person with suicidal behaviour [26,30,31,35,38,41,45,47,51,54]. The study by Miche et al. [41] found that ridge regression performed better than lasso regression in predicting suicide attempts in adolescents and young adults (AUC = 0.829). Hettige et al. [30] discovered that the Elastic net had lower accuracy (0.65) compared to Lasso regression (0.67), but similar results for AUC (0.71) in classifying individuals as suicide attempters or non-suicide attempters.

We identified twenty-six studies that used the ensemble method to classify and predict suicidal behaviours. The ensemble method is one of the machine learning techniques that is currently and widely used in healthcare [59,60]. The same goes for the study of suicidal behaviour prediction, where most of the studies identified in this systematic review apply and utilize the ensemble method for predicting suicidal behaviour

[8,50,54–57]. The ensemble method trains multiple learners to solve the same problem [61]. An ensemble contains a number of learners called base learners. Base learners are usually generated from training data by a base learning algorithm (decision tree, neural network).

Three common models for ensemble methods studied in this systematic review are bagging, voting, random forest and boosting. Bagging is shorthand for combining bootstrapping and aggregation, combines the base learners in a parallel manner on different bootstrap samples and then aggregates the individual predictions with average weight to form a final prediction [62]. Voting is the simplest way to combine base learners to produce a final prediction based on a majority vote (hard voting) or average sum of predicted probabilities (soft voting) in the class label. In the study by Nordin et al. [50], bagging and voting models were used to predict suicide attempts and found that both models achieved the highest accuracy of 0.92.

Random Forest is the state-of-the-art ensemble method, and it is an extension of bagging. The main difference is the incorporation of randomized feature selection. When constructing a large number of decision trees, Random Forest first randomly selects a subset of features at each split selection step and then performs the usual split selection procedure within the selected subset of features. The analysis shows that the majority of studies (26 studies) currently utilized Random Forest as a machine learning technique for predicting suicidal behaviour. Kim et al. [57] compare Random Forest with k-nearest neighbours for detecting suicide risk in college students and the results show that random forest achieves high performance (AUC = 0.851, Precision = 0.953). Navarro et al. [56] proposed a suicide attempts model for the young population using Random Forest while Cho et al. [55] proposed the same technique for suicide deaths among the elderly population. Both studies highlighted the good performance of both Random Forest models for specificity (0.76–0.833) and AUC (0.76–0.818).

Another ensemble method is known as boosting, in which the model creates a strong model based on multiple weaker models. In boosting, a weak base learner is first created and then, the accuracy of the model is improved by sequentially adding more weak base learners [61,62]. The collection of weak base learners forms a robust classification model. Two common methods of boosting are AdaBoost and Gradient Boosting. AdaBoost is an adaptive boosting in which the weights for the base learners are assigned based on the accuracy of the base learner, and the weights of the training data are changed based on the accuracy of the prediction, while Gradient Boosting is a powerful algorithm for building predictive models because it provides more accurate results and performs optimization in the function space, which facilitates the use of custom loss function easier [63]. We identified eight studies that applied the boosting model to predict suicidal behaviour [27,31,36,43,45,46,48,53]. Most of the identified studies showed that boosting models are able to classify a person with suicidal behaviour and achieve the highest performance (accuracy = 0.76–0.90, sensitivity = 0.75–0.89, AUC = 0.80–0.86) compared to the single prediction models (decision tree, Naïve Bayes, logistic regression). The overall review of the identified studies using types of machine learning techniques for predicting suicidal behaviour is shown in Fig. 4.

3.3. Categories of variables/factors/indicators/predictors for suicidal behaviour prediction

Our studies identified ten (10) major categories of factors in developing models using machine learning techniques for suicidal behaviour, as shown in Table 4. The classification categories for the factors are based on the studies by [2,14].

The majority of studies (n = 30) used demographic features as indicators in developing their models to predict suicidal behaviour, including age, gender, race, ethnicity, marital status, employment, education, and socioeconomic status. Demographic features are known as the most important predictors for classifying suicidal behaviours [2]. Most studies considered different ages (adolescents, young, adults), gender (females, males), race (Black, Asian, White, Alaskan), marital status (married, single, widowed), employment (employed, unemployed), and education (elementary school, middle school, high school, university). In the study by Ryu et al. [34], demographic features were used as characteristics to classify suicide ideators and non-suicide ideators. The results showed that age, gender, education, and employment were the most important features for the prediction model using random forest.

Two studies used military features such as a number of deployments, a branch of service (navy, marines, army, air force), total time deployed,

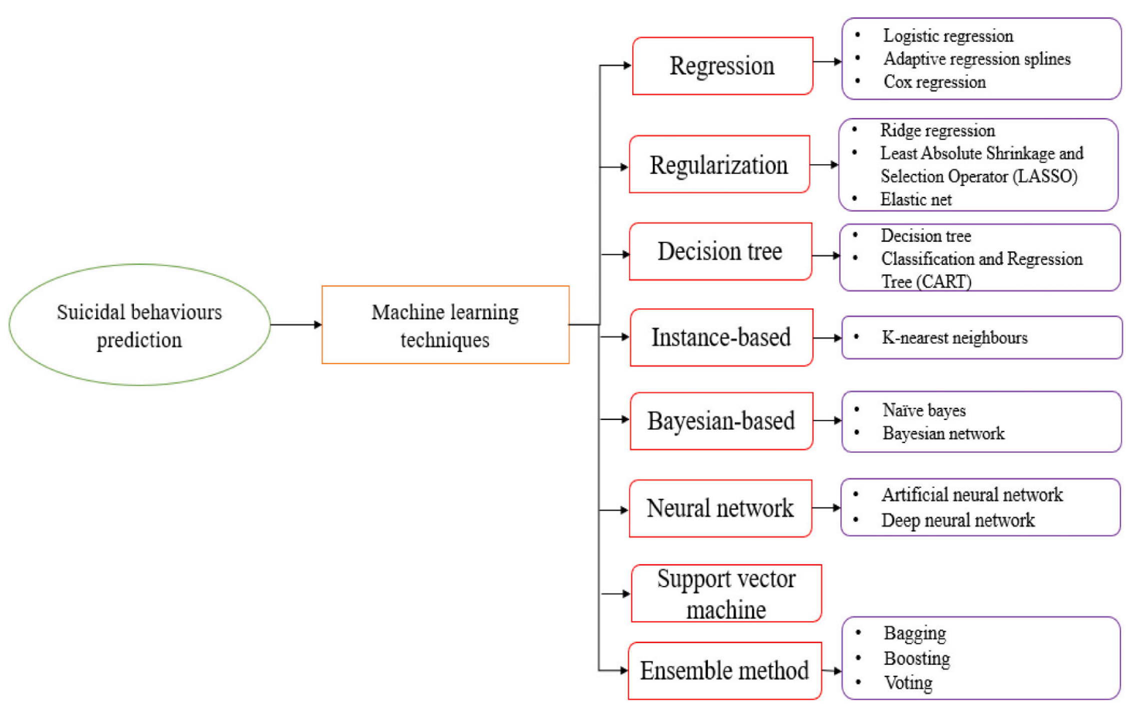

Fig. 4. Classification of machine learning techniques used for suicidal behaviour prediction.

Table 4
Categories of factors as an input for machine learning techniques.

No.	Categories	Studies
1.	Demographics (e.g.: age, gender, race, ethnicity, marital status, employment, education, socio-economic status)	[8,27,29,30,33,35,37,44,47,49–51,54–56]
2.	Military characteristics (e.g.: number of deployment)	[29,31]
3.	Family history of psychopathology or suicidal behaviours or disease (e.g.: parents' psychological state, parenting practices, maternal depression, relative suicide attempt)	[30,32,41,45,47,50,52,56]
4.	Physical illness/health and physical characteristics (e.g.: weight, height, BMI, physical disability, systolic blood pressure, waist measurement, asthma, cancer)	[8,28,33,34,36,40,47,52,54,55]
5.	Treatment history (e.g.: specific medication use, psychiatric drug use, prior psychiatric hospitalization)	[5,8,26,30,31,35,38,40,41,46,47,49–51,55]
6.	Internalizing psychopathology (e.g.: mood disorder, anxiety disorder, depressive disorder, emotion dysregulation)	[8,18,26–28,32,34,35,37–46,50–54,57]
7.	Externalizing psychopathology (e.g.: substance use, aggressive behaviour, impulsivity, antisocial behaviour)	[8,30,32,35,36,39,46,48,50–53,57]
8.	Social factors (e.g.: abuse history, alcoholic history, stressful life events, family problems, peer problems)	[5,26–28,30,32,39–41,44,50,52,54,57]
9.	Prior suicidal behaviours/ thoughts (e.g.: suicide attempt, suicide ideation, suicide plan)	[5,26,30,32,35–39,41,42,44,45,47,50,52]
10.	Cognitive abilities/problems (e.g.: cognitive difficulties, mental state, intelligence, school performance, neurological problem)	[27,32,34,41,47,52,53,57]

and rank (officer, enlisted) as inputs for predictors in developing predictive models of suicidal behaviour among veterans [29,31,43]. In addition, eight studies used predictors from a family history of psychopathology factors. Family history of psychopathology features included maternal depression, familial alcoholism, history of suicide and suicide attempts among family members [30,32,41,45,47,50,52,56]. Of the eight studies, only three studies [30,32,45] ranked family history of psychopathology as one of the most important predictors for predicting suicidal behaviours.

In addition, ten studies were identified that used physical illness feature as one of the predictors of suicidal behaviour [8,28,33,34,36,40,47,52,54,55]. Physical illness also refers to health comorbidities, and physical characteristics such as handicap, weight, height, body mass index (BMI), waist circumference, cancer, and heart disease. Five studies indicate that physical illnesses are closely related to suicidal behaviour and may be important features that distinguish suicide ideators and suicide attempters [33,34,36,40,55].

Based on the systematic review, fifteen studies were identified that used predictor categories of treatment history [5,35,49,51,55]. In general, treatment history for suicidal behaviours includes prior psychiatric hospitalizations, use of specific medication and psychiatric drug use. Iorfino et al. [47] identified treatment with antipsychotics as one of the predictors of self-harm in adolescents, and the studies by Passos et al. [26] and Horvath et al. [46] also highlighted that psychiatric hospitalization is an important feature in predicting suicidal behaviour. Furthermore, the number of hospitalizations was a significant and meaningful feature associated with suicide risk in schizophrenia [30].

Internalizing psychopathology variables were identified in 24 studies as the most popular predictor categories for predicting suicidal behaviour in an individual. Internalizing psychopathology refers to a condition characterized by negative emotions within the self, including mood disorder, depressive disorder, anxiety disorder, and emotion dysregulation [64]. The presence of internalizing psychopathology in predicting suicidal behaviour is critical, as most studies have used these features in their predictive models [41,43,44,46,54]. However, only a few studies found that internalizing psychopathology, especially depressive disorder, was the most significant features in predicting suicide attempts [37–39,44,46,50].

Externalizing psychopathology is one of the predictor categories for suicidal behaviour. Externalizing psychopathology categories refer to a variety of co-occurring psychiatric disorders in which the actions occur primarily in the external world, such as aggressiveness, substance use disorders, antisocial personality, incarceration history, and impulsivity [64]. Thirteen studies were identified using externalizing psychopathology features as a predictor of suicidal behaviour [8,30,32,35,36,39,46,48,50–53,57]. Substance use disorder was identified as the most important predictor in the externalizing psychopathology category for predicting suicidal behaviour [35,36,50].

Fourteen studies were identified in which social factors were used as predictors of suicidal behaviours. Many social factors/features have been used as a predictor to predict suicidal behaviours in an individual including family problems, peer problems, stressful life events, abuse history and alcoholic history [27,30,32,44]. Alcoholic history is one of the highest features importance in the study by Shen et al. [39] and Nordin et al. [50]. Cannabis, cocaine, and substance abuse/dependence have also been identified as crucial predictors for predicting suicidal behaviour [26,30]. The studies by Barros et al. [27] and Miche et al. [41] found that social factors/features were the important predictors in classifying and predicting suicidal behaviour.

Past suicidal behaviours and thoughts were identified as the second most important categories after internalizing psychopathology features. Sixteen studies used and applied the presence of suicidal behaviours in an individual's past to predict suicidal behaviours in the future [32,35,37,47,52]. Most studies found that individuals with a history of suicidal behaviour, including suicidal thoughts, suicidal ideation, and suicidal attempts were at the highest risk for suicide [8,26,36,37]. Studies by Burke et al. [44], Miche et al. [41], and Nordin et al. [50] highlighted that suicide attempt history is the most important and significant feature for predicting suicidal behaviour.

Cognitive abilities, also known as cognitive problems, where individuals have difficulties in the areas of cognition, intelligence, and mental state [2,14]. Eight studies were identified that used cognitive ability features to predict suicidal behaviours [27,32,34,41,47,52,53,57]. Recently, a study by Kim et al. [57] used cognitive scales as one of the clinical assessment scores to detect suicide risk. McMullen et al. [53] also used the feature of loss of cognitive control as one of the criteria for predicting suicidal ideation in high-risk inpatients. The presence of a neurological problem is also one of the predictors used in the study by Iorfino et al. [47]. In addition, Fig. 5 shows the importance of risk factors based on existing studies used in this review to develop predictive models using the ranking score method.

4. Discussion

This systematic review provides a summary of studies using machine learning techniques to improve the understanding and prediction of suicidal behaviour. The current studies included results from 35 articles,

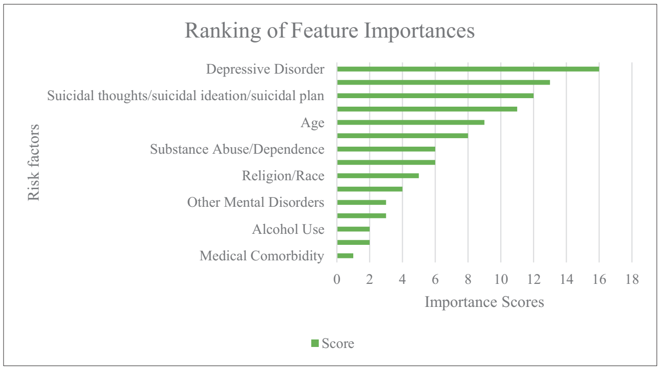

Fig. 5. Ranking of the importance of risk factors based on existing studies for predicting suicidal behaviour.

all published within the last six years. As shown in Fig. 6, most studies have taken several steps to develop models to predict suicidal behaviour. These include the collection of data from various sources (clinical research databases and electronic health records), pre-processing of the data (e.g.: imputation of missing values, normalization), feature selection (dimensional reduction), machine learning models selection, and evaluation of the models. Based on these findings, we determine that machine learning techniques have been shown to be able to predict suicidal behaviours. Although we have recently observed an increasing use of machine learning techniques, we conclude that these techniques are limited and have little application in the study of suicidal behaviour prediction. In this section, our goal is to identify gaps in this literature and to highlight challenges and opportunities in developing predictive models of suicidal behaviour using machine learning techniques.

The main challenge in developing predictive models of suicidal behaviour using machine learning techniques is imbalanced data classification [28,32–34,44,47,56]. Imbalanced data classification is defined when the number of samples belonging to one class is significantly lower than the number of samples belonging to other classes [60]. Generally, the classes with more samples are called majority classes, while the classes with fewer samples are minority classes. An imbalance in class distribution is more challenging to achieve accurate performance and requires specialized techniques. The imbalanced class distribution was found in the prediction of suicidal behaviour prediction due to the low frequency of suicide attempts and a lower number of positive cases in the samples [53] and resulting in a large number of false positives [49]. We found that current studies addressed the problem of imbalanced classes only to a limited extent. Only the study by Oh et al. [40] used an oversampling technique to rebalance the data; however, this technique increases the likelihood of overfitting because it replicates the minority class (suicide attempters). Therefore, there is a need and opportunity to continue research on imbalanced classification for predicting suicidal behaviour, which could be practical and useful for clinicians in effectively classifying a person with suicidal behaviour.

Overfitting is another challenge in developing predictive models for suicidal behaviour. Machine learning involves strategies to ensure robustness against overfitting, for example, when a model is very specific to a training dataset but fails when applied to new datasets. Overfitting is more expected when the model is overly complex, or the number of features is very large, but the sample size of the dataset is small [13]. Based on this systematic review, overfitting occurs in some studies when too many predictors and a small sample size are used [35,44]. Some techniques to protect against overfitting include regularization (synthetically requiring smoothness in the model) and early stopping (stopping iterations when a certain performance level is reached) [8,38,41]. Although many studies also employ cross-validation techniques (such as stratified sampling, hold-out sampling, bootstrapping sampling), few studies have validated the models on external samples [30,40]. Thus, replication of a model trained on one dataset to another dataset developed from a similar focus population is needed and generalization to other samples may be necessary in future work to reproduce and extend the current results in a different setting [9,49,51].

Missing value problem is important because it affects the performance of machine learning models. In the study [24], the presence of missing values affects the evaluation of the model developed and an attempt is made to minimize the missing values by using an expectation-maximization algorithm to impute the missing values. In the study of [53], it was found that they suffered from missing value and loss of suicide cases after merging the data due to missing demographic information. Therefore, when developing predictive models using machine learning algorithms, attention must be paid to the missing data, as missing values could affect the correlation between features in the dataset and the performance of the models [60].

The majority of studies describe machine learning classification tasks that aim to identify and classify whether or not a person is exhibiting suicidal behaviour [27,32,37,47,57]. In general, predictive models have been developed for binary classification tasks where researchers attempt to make predictions for two classes (non-suicidal attempters or suicidal attempters). However, considering a mental illness such as suicidality as a single category may not take into account the variability in risk level and how the illness reveals [17]. In everyday psychiatric practice, it is often argued that the more difficult issue is often not detecting the presence of suicidal behaviour conditions, but the need to monitor risk levels for suicidal behaviour or response to treatment, rather than psychiatric case specificity (positive case or negative case). This allows for more accurate decisions about further management in the treatment of suicidal behaviour [6]. Therefore, multi-task learning or multi-class prediction may be appropriate to use machine learning techniques to examine differences between risk levels for suicidal behaviour and treatment groups, and to identify important and undetected features.

A variety of machine learning techniques have been applied in a number of studies. Ensemble methods, which combine predictions from a variety of models rather than using just one, provide better predictions

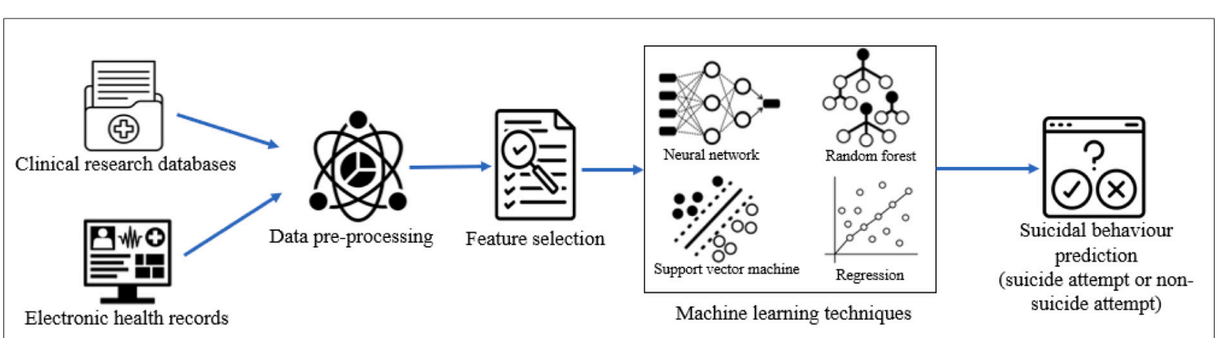

Fig. 6. Current overview of suicidal behaviour prediction models using machine learning.

than single models. Most studies have demonstrated that the use of ensemble methods can improve the accuracy of models [27,34,44,46]. However, ensemble methods suffer from the lack of interpretability and explainability of the predictions [13,50]. In the study by Oh et al. [32], it was highlighted that the inner workings of machine learning techniques act as a 'black box', which makes it more difficult to interpret the meaning of the models created. This is one of the main challenges for machine learning where the predictions cannot be explained and interpreted [65]. In the healthcare domain, the issue of transparency is related to machine learning and the lack of trust in the models creates the need for predictive models that can be explained [66]. This is where explainable artificial intelligence/machine learning (XAI) comes into the picture. Explainable models (transparent techniques and opaque techniques) are very important to facilitate the understanding of various aspects of a model to increase the transparency and robustness of the model [67]. Therefore, it is important to understand how well the machine learning techniques translate from training data to an individual healthcare system especially in predicting suicide attempters. Future studies will therefore need to explore how best to present, visualize and communicate the results of these explainable models for an individual with suicidal behaviour, so that they are beneficial, intuitive, and well-known to the clinician and patient. The merging of machine learning predictions with clinician-based suicide risk assessments needs to be further investigated in order to develop effective decision support tools for clinicians.

Moreover, Natural Language Processing (NLP) techniques for predicting suicidal behaviour have the potential to understand posts and texts on social media [12]. This is because pattern recognition and sentiment analysis can identify posts that contain suicide-related material and provide resources to help the person who posted the content on social media [68]. However, predicting suicidal behaviour using social media analytics should be done carefully, because each person with suicidal behaviour has different risk factors than others, and may have potentially harmful content. Suicide-related material posted on social media must be classified appropriately, as the posts could also be fabricated and false [10]. Further evaluations of safety and effectiveness are also needed for suicide prevention strategies that use mobile applications. In addition to applying text mining with NLP techniques, recent studies have focused on detecting suicidal behaviour using clinical notes from electronic health records [69,70]. The unstructured clinical notes in electronic health records can help improve predictive performance in classifying individuals with suicidal behaviour without relying solely on structured codes. The potential of this data structure is highlighted in the recent studies through the use of clinical notes that provide innovative results [69,70]. Therefore, a future area of research is needed for detection and prediction of suicidal behaviour using clinical notes from electronic health records to assist clinicians in suicide prevention efforts and accurate clinical decision making.

Our systematic review has several limitations. First, we discovered only thirty-five empirical studies that met our inclusion criteria, and the included studies focused only on electronic health records and clinical research data. This paper did not examine the effectiveness of machine learning techniques in social media data and unstructured data. Second, this paper addresses the prediction of suicidal behaviour based on supervised learning studies (classification) and does not investigate studies related to unsupervised learning studies (clustering, association). Third, this paper does not report the pre-processing techniques used to develop predictive models of suicidal behaviour. Pre-processing techniques are important steps in the development of predictive models. Therefore, further investigation could explore these pre-processing techniques for predicting suicidal behaviour. Fourth, the features (risk factors) discussed in this review are general and related to all suicidal behaviours in predicting suicidal behaviour. Future research attempts to specify and discuss risk factors for specific suicidal behaviours.

5. Conclusion

The purpose of this review was to evaluate and analyze the state-of-the-art of machine learning techniques for predicting suicidal behaviour. A number of different types of machine learning techniques have been proposed to develop a prediction model for suicidal behaviour. However, due to the complexity and dynamic characteristics of suicidal behaviours, it remains difficult to develop a universal prediction model that provides accurate prediction. To conclude, machine learning research in suicidal behaviour has made exciting progress recently. Machine learning has been used to enhance the decision-making process and provide better detection to overcome the problem of undetected and misdiagnosed individuals with suicidal behaviours. Machine learning has also shown to have a promising future given the increasing availability of electronic health records and clinical research data. Future research needs to address persistent methodological problems by integrating novel techniques to address imbalanced data classification, overfitting, missing values, and classification tasks. However, machine learning technique has general limitations in interpretability and explainability. Therefore, expanding to improve the interpretability and explainability of machine learning-based predictive models is required and clinically appropriate performance metrics would be necessary to translate these models for use in everyday clinical practice. As machine learning techniques become more accessible to clinicians and researchers, it is expected that the field of suicidality will continue to grow and prevention efforts to reduce suicide rates will be successful and have a positive impact.

Declaration of competing interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgement

The authors would like to thank all parties that support and involve directly or indirectly into this research, especially Universiti Sains Malaysia, Pulau Pinang and Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur. This research is supported by the Ministry of Higher Education, Malaysia for Fundamental Research Grant Scheme (FRGS) with Project Code: FRGS/1/2020/ICT02/USM/02/5.

References

[1] World Health Organization. Suicide. https://www.who.int/news-room/fact-sheets/detail/suicide. [Accessed 20 August 2021].
[2] Franklin JC, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017;143(2):187–232. https://doi.org/10.1037/bul0000084.
[3] O'Connor RC, Nock MK. The psychology of suicidal behaviour. Lancet Psychiatry Jun. 2014;1(1):73–85. https://doi.org/10.1016/S2215-0366(14)70222-6.
[4] Turecki G, et al. Suicide and suicide risk. Nat Rev Dis Primers Dec. 2019;5(1):74. https://doi.org/10.1038/s41572-019-0121-0.
[5] Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. Clin Psychol Sci May 2017;5(3):457–69. https://doi.org/10.1177/2167702617691560.
[6] Velupillai S, et al. Risk assessment tools and data-driven approaches for predicting and preventing suicidal behavior. Front Psychiatry Feb. 2019;10:36. https://doi.org/10.3389/fpsyt.2019.00036.
[7] Ribeiro JD, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol Med Jan. 2016;46(2):225–36. https://doi.org/10.1017/S0033291715001804.

[8] Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. J Child Psychol Psychiatry Dec. 2018;59(12):1261–70. https://doi.org/10.1111/jcpp.12916.
[9] Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol May 2018;14(1):91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037.
[10] Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health Aug. 2020;17(16):5929. https://doi.org/10.3390/ijerph17165929.
[11] Fonseka TM, Bhat V, Kennedy SH. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. Aust N Z J Psychiatry Oct. 2019;53(10):954–64. https://doi.org/10.1177/0004867419864428.
[12] Alpaydin E. Introduction to machine learning. 2nd ed. Cambridge, Mass: MIT Press; 2010.
[13] Boudreaux ED, et al. Applying machine learning approaches to suicide prediction using healthcare data: overview and future directions. Front Psychiatry Aug. 2021; 12:707916. https://doi.org/10.3389/fpsyt.2021.707916.
[14] Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review. J Affect Disord Feb. 2019;245:869–84. https://doi.org/10.1016/j.jad.2018.11.073.
[15] Alonso SG, et al. Data mining algorithms and techniques in mental health: a systematic review. J Med Syst Sep. 2018;42(9):161. https://doi.org/10.1007/s10916-018-1018-2.
[16] Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med Jul. 2019;49(09): 1426–48. https://doi.org/10.1017/S0033291719000151.
[17] Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput-Hum Interact Oct. 2020;27(5): 1–53. https://doi.org/10.1145/3398069.
[18] Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR. Suicide prediction models: a critical review of recent research with recommendations for the way forward. Mol Psychiatry Jan. 2020;25(1):168–79. https://doi.org/10.1038/s41380-019-0531-0.
[19] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med Jul. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097.
[20] Hutton B, Catalá-López F, Moher D. The PRISMA statement extension for systematic reviews incorporating network meta-analysis: PRISMA-NMA. Med Clín Engl Ed Sep. 2016;147(6):262–6. https://doi.org/10.1016/j.medcle.2016.10.003.
[21] Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. Acta Psychiatr Scand Jun. 2009;119(6):443–50. https://doi.org/10.1111/j.1600-0447.2009.01388.x.
[22] PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev Dec. 2015;4(1):1. https://doi.org/10.1186/2046-4053-4-1.
[23] Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev Dec. 2017;6(1):245. https://doi.org/10.1186/s13643-017-0644-y.
[24] De Leo D, et al. International study of definitions of english-language terms for suicidal behaviours: a survey exploring preferred terminology. BMJ Open Feb. 2021;11(2):e043409. https://doi.org/10.1136/bmjopen-2020-043409.
[25] Amini P, Ahmadinia H, Poorolajal J, Amiri MMoqaddasi. Evaluating the high risk groups for suicide: a comparison of logistic regression, support vector machine, decision tree and artificial neural network. Iran J Public Health Sep. 2016;45(9): 1179–87.
[26] Passos IC, et al. Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J Affect Disord Mar. 2016;193:109–16. https://doi.org/10.1016/j.jad.2015.12.066.
[27] Barros J, et al. Suicide detection in Chile: proposing a predictive model for suicide risk in a clinical sample of patients with mood disorders. Rev Bras Psiquiatr Oct. 2016;39(1):1–11. https://doi.org/10.1590/1516-4446-2015-1877.
[28] Barak-Corren Y, et al. Predicting suicidal behavior from longitudinal electronic health records. Am J Psychiatry Feb. 2017;174(2):154–62. https://doi.org/10.1176/appi.ajp.2016.16010077.
[29] Gradus JL, King MW, Galatzer-Levy I, Street AE. Gender differences in machine learning models of trauma and suicidal ideation in veterans of the Iraq and Afghanistan wars: gender differences in trauma and suicidal ideation. J Trauma Stress Aug. 2017;30(4):362–71. https://doi.org/10.1002/jts.22210.
[30] Hettige NC, et al. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: a machine learning approach. Gen Hosp Psychiatry Jul. 2017;47:20–8. https://doi.org/10.1016/j.genhosppsych.2017.03.001.
[31] Kessler RC, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the veterans health administration. Int J Methods Psychiatr Res 2017;26(3):Sep. https://doi.org/10.1002/mpr.1575.
[32] Oh J, Yun K, Hwang J-H, Chae J-H. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front Psychiatry Sep. 2017;8:192. https://doi.org/10.3389/fpsyt.2017.00192.
[33] Choi SB, Lee W, Yoon J-H, Won J-U, Kim DW. Ten-year prediction of suicide death using cox regression and machine learning in a nationwide retrospective cohort study in South Korea. J Affect Disord Apr. 2018;231:8–14. https://doi.org/10.1016/j.jad.2018.01.019.
[34] Ryu S, Lee H, Lee D-K, Park K. Use of a machine learning algorithm to predict individuals with suicide ideation in the general population. Psychiatry Investig Nov. 2018;15(11):1030–6. https://doi.org/10.30773/pi.2018.08.27.
[35] Simon GE, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. Am J Psychiatry Oct. 2018;175 (10):951–60. https://doi.org/10.1176/appi.ajp.2018.17101167.
[36] Jung JS, Park SJ, Kim EY, Na K-S, Kim YJ, Kim KG. Prediction models for high risk of suicide in korean adolescents using machine learning techniques. PLOS ONE Jun. 2019;14(6):e0217639. https://doi.org/10.1371/journal.pone.0217639.
[37] Ribeiro JD, Huang X, Fox KR, Walsh CG, Linthicum KP. Predicting imminent suicidal thoughts and nonfatal attempts: the role of complexity. Clin Psychol Sci Sep. 2019;7(5):941–57. https://doi.org/10.1177/2167702619838464.
[38] Su C, Aseltine R, Doshi R, Chen K, Rogers SC, Wang F. Machine learning for suicide risk prediction in children and adolescents with electronic health records. Transl Psychiatry Dec. 2020;10(1):413. https://doi.org/10.1038/s41398-020-01100-0.
[39] Shen Y, et al. Detecting risk of suicide attempts among chinese medical college students using a machine learning algorithm. J Affect Disord Aug. 2020;273: 18–23. https://doi.org/10.1016/j.jad.2020.04.057.
[40] Oh B, Yun J-Y, Yeo EC, Kim D-H, Kim J, Cho B-J. Prediction of suicidal ideation among korean adults using machine learning: a cross-sectional study. Psychiatry Investig Apr. 2020;17(4):331–40. https://doi.org/10.30773/pi.2019.0270.
[41] Miché M, et al. Prospective prediction of suicide attempts in community adolescents and young adults, using regression methods and machine learning. J Affect Disord Mar. 2020;265:570–8. https://doi.org/10.1016/j.jad.2019.11.093.
[42] van Mens K, et al. Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study. J Affect Disord Jun. 2020;271:169–77. https://doi.org/10.1016/j.jad.2020.03.081.
[43] Lin G-M, Nagamine M, Yang S-N, Tai Y-M, Lin C, Sato H. Machine learning based suicide ideation prediction for military personnel. IEEE J Biomed Health Inform Jul. 2020;24(7):1907–16. https://doi.org/10.1109/JBHI.2020.2988393.
[44] Burke TA, Jacobucci R, Ammerman BA, Alloy LB, Diamond G. Using machine learning to classify suicide attempt history among youth in medical care settings. J Affect Disord May 2020;268:206–14. https://doi.org/10.1016/j.jad.2020.02.048.
[45] Chen Q, et al. Predicting suicide attempt or suicide death following a visit to psychiatric specialty care: a machine learning study using swedish national registry data. PLoS Med Nov. 2020;17(11):e1003416. https://doi.org/10.1371/journal.pmed.1003416.
[46] Horvath A, Dras M, Lai CCW, Boag S. Predicting suicidal behavior without asking about suicidal ideation: machine learning and the role of borderline personality disorder criteria. Suicide Life Threat Behav Jun. 2021;51(3):455–66. https://doi.org/10.1111/sltb.12719.
[47] Iorfino F, et al. Predicting self-harm within six months after initial presentation to youth mental health services: a machine learning study. PLOS ONE Dec. 2020;15 (12):e0243467. https://doi.org/10.1371/journal.pone.0243467.
[48] van Mens K, et al. Applying machine learning on health record data from general practitioners to predict suicidality. Internet Interv Sep. 2020;21:100337. https://doi.org/10.1016/j.invent.2020.100337.
[49] Edgcomb JB, Shaddox T, Hellemann G, Brooks JO. Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness. J Psychiatr Res Apr. 2021;136:515–21. https://doi.org/10.1016/j.jpsychires.2020.10.024.
[50] Nordin N, Zainol Z, Noor MHMohd, Fong CLai. A comparative study of machine learning techniques for suicide attempts predictive model. Health Informatics J Jan. 2021;27(1). https://doi.org/10.1177/1460458221989395. p. 146045822198939.
[51] Kirlic N. A machine learning analysis of risk and protective factors of suicidal thoughts and behaviors in college students. J Am Coll Health Jul. 2021:1–10. https://doi.org/10.1080/07448481.2021.1947841.
[52] Macalli M, et al. A machine learning approach for predicting suicidal thoughts and behaviours among college students. Sci Rep Dec. 2021;11(1):11363. https://doi.org/10.1038/s41598-021-90728-z.
[53] McMullen L, Parghi N, Rogers ML, Yao H, Bloch-Elkouby S, Galynker I. The role of suicide ideation in assessing near-term suicide risk: a machine learning approach. Psychiatry Res Oct. 2021;304:114118. https://doi.org/10.1016/j.psychres.2021.114118.
[54] van Vuuren CL, et al. Comparing machine learning to a rule-based approach for predicting suicidal behavior among adolescents: results from a longitudinal population-based survey. J Affect Disord Dec. 2021;295:1415–20. https://doi.org/10.1016/j.jad.2021.09.018.
[55] Cho S-E, Geem ZW, Na K-S. Development of a suicide prediction model for the elderly using health screening data. Int J Environ Res Public Health Sep. 2021;18 (19):10150. https://doi.org/10.3390/ijerph181910150.
[56] Navarro MC, et al. Machine learning assessment of early life factors predicting suicide attempt in adolescence or young adulthood. JAMA Netw Open Mar. 2021;4 (3):e211450. https://doi.org/10.1001/jamanetworkopen.2021.1450.
[57] Kim S, Lee H-K, Lee K. Detecting suicidal risk using MMPI-2 based on machine learning algorithm. Sci Rep Dec. 2021;11(1):15310. https://doi.org/10.1038/s41598-021-94839-5.
[58] Faul AC. A concise introduction to machine learning. Boca Raton, Florida: CRC Press; 2019.

[59] Woldaregay AZ, et al. Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. Artif Intell Med Jul. 2019;98:109–34. https://doi.org/10.1016/j.artmed.2019.07.007.

[60] Parimbelli E, et al. A review of AI and data science support for cancer management. Artif Intell Med Jul. 2021;117:102111. https://doi.org/10.1016/j.artmed.2021.102111.

[61] Zhou Z-H. Ensemble methods: foundations and algorithms. 1st ed. Chapman and Hall/CRC; 2012. https://doi.org/10.1201/b12207.

[62] Kantardzic M. DATA MINING: concepts, models, methods, and algorithms. Place of publication not identified. JOHN WILEY; 2019.

[63] Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O. Predictive analytics with gradient boosting in clinical medicine. Ann Transl Med Apr. 2019;7(7). https://doi.org/10.21037/atm.2019.03.29. 152-152.

[64] Krueger RF, Markon KE. Understanding psychopathology: melding behavior genetics, personality, and quantitative psychology to develop an empirically based model. Curr Dir Psychol Sci 2006;15(3):113–7. https://doi.org/10.1111/j.0963-7214.2006.00418.x.

[65] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak Dec. 2020;20(1):310. https://doi.org/10.1186/s12911-020-01332-6.

[66] Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning-based prediction models in healthcare. WIREs Data Min Knowl Discovery Sep. 2020;10(5). https://doi.org/10.1002/widm.1379.

[67] Belle V, Papantonis I. Principles and practice of explainable machine learning. Front Big Data Jul. 2021;4:688969. https://doi.org/10.3389/fdata.2021.688969.

[68] Roy A, Nikolitch K, McGinn R, Jinah S, Klement W, Kaminsky ZA. A machine learning approach predicts future risk to suicidal ideation from social media data. Npj Digit Med Dec. 2020;3(1):78. https://doi.org/10.1038/s41746-020-0287-6.

[69] Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep Dec. 2018;8(1):7426. https://doi.org/10.1038/s41598-018-25773-2.

[70] Cusick M, et al. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. J Psychiatr Res Apr. 2021;136:95–102. https://doi.org/10.1016/j.jpsychires.2021.01.052.