Journal of Affective Disorders 245 (2019) 869–884

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Review article

The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review

Taylor A. Burke[a,1,*], Brooke A. Ammerman[b,1], Ross Jacobucci[b]

[a] Temple University, Department of Psychology, Philadelphia, PA, USA
[b] University of Notre Dame, Department of Psychology, Notre Dame, IN, USA

ARTICLE INFO

Keywords:
Machine learning
Suicide
Suicide attempt
Suicide risk
Suicidal ideation
Non-suicidal self-injury
Big data
Pattern recognition
Exploratory data mining

ABSTRACT

Background: Machine learning techniques offer promise to improve suicide risk prediction. In the current systematic review, we aimed to review the existing literature on the application of machine learning techniques to predict self-injurious thoughts and behaviors (SITBs).

Method: We systematically searched PsycINFO, PsycARTICLES, ERIC, CINAHL, and MEDLINE for articles published through February 2018.

Results: Thirty-five articles met criteria to be included in the review. Included articles were reviewed by outcome: suicide death, suicide attempt, suicide plan, suicidal ideation, suicide risk, and non-suicidal self-injury. We observed three general aims in the use of SITB-focused machine learning analyses: (1) improving prediction accuracy, (2) identifying important model indicators (i.e., variable selection) and indicator interactions, and (3) modeling underlying subgroups. For studies with the aim of boosting predictive accuracy, we observed greater prediction accuracy of SITBs than in previous studies using traditional statistical methods. Studies using machine learning for variable selection purposes have both replicated findings of well-known SITB risk factors and identified novel variables that may augment model performance. Finally, some of these studies have allowed for subgroup identification, which in turn has helped to inform clinical cutoffs.

Limitations: Limitations of the current review include relatively low paper sample size, inconsistent reporting procedures resulting in an inability to compare model accuracy across studies, and lack of model validation on external samples.

Conclusions: We concluded that leveraging machine learning techniques to further predictive accuracy and identify novel indicators will aid in the prediction and prevention of suicide.

1. Introduction

Suicide is a major public health problem, with an estimated 800,000 deaths as a result of suicide each year (WHO, 2014), ranking it the second leading cause of death among ages 10–34 (Center for Disease Control and Prevention [CDC], 2016). Nonfatal suicide attempts (SA; a non-fatal, self-directed, potentially injurious behavior with an intent to die as a result of the behavior), suicidal planning (SP; formulation of a specific method through which one intends to die), suicidal ideation (SI; thinking about or considering suicide) and non-suicidal self-injury (NSSI; self-directed injurious behavior without an intent to die as a result of the behavior) are also of significant concern given that they are associated with substantial personal, family and economic burden (e.g., Crosby et al., 2011; Nock, 2009; Shepard et al., 2016). Moreover, previous SAs, SP, SI, and NSSI represent some of the strongest predictors of future suicidal behavior (Ribeiro et al., 2016) and even death by suicide (Shepard et al., 2016). Although a large body of research has aimed to identify those at risk for suicide and nonfatal SAs, a recent large meta-analysis suggests that our ability to predict these behaviors has remained limited over the past five decades (Franklin et al., 2017). Similarly, the prediction of NSSI has continued to be weak (Fox et al., 2015). In part, the field has been limited by an over-reliance on traditional statistical approaches (Franklin et al., 2017), which often restricts the number of variables that can be simultaneously examined, thus forcing researchers to use overly simplistic models for prediction. Given the complexity of suicide, such approaches have hampered our ability to inform clinical decision making in a meaningful way (e.g., Curtin et al., 2016; Walsh, Ribeiro, & Franklin, 2017).

* Corresponding author.
E-mail address: taylor.burke@temple.edu (T.A. Burke).
[1] Taylor A. Burke and Brooke A. Ammerman contributed equally towards first authorship for this study.

https://doi.org/10.1016/j.jad.2018.11.073
Received 10 August 2018; Received in revised form 20 October 2018; Accepted 11 November 2018
Available online 12 November 2018
0165-0327/

To help address the limitations of traditional statistical approaches, psychological research has recently begun utilizing machine learning (ML) techniques (McArdle & Ritschard, 2014). Traditional approaches constrain predictive accuracy in several important ways that are attended to in ML methods. For example, traditional approaches greatly minimize the number of predictors and interactions that can be examined simultaneously and impose linearity on relationships that likely have more complex associations (McArdle & Ritschard, 2014; Morgan, 2005). On the other hand, ML approaches allow for the simultaneous testing of numerous factors and their complex interactions (McArdle & Ritschard, 2014). They also allow for non-linearity in producing predictive models. Furthermore, traditional statistical methods rely on the researcher to define the relation between predictors and outcomes a priori. In turn, this prevents the consideration of innumerable pathways likely present in the prediction of complex psychological phenomena (i.e., suicide-related events). Instead, ML techniques are able to iteratively test all possible relationships and identify the superlative set of algorithm operations to augment accuracy (McArdle & Ritschard, 2014).

2. Current study

The advantages of ML approaches have the potential to significantly impact prediction of suicide-related events, and thus, improve suicide prevention and intervention efforts. Given the recent popularity of ML implementation, the aim of the current study was to conduct a systematic review of empirical articles employing ML techniques to improve prediction and classification of suicidal and/or non-suicidal self-injurious thoughts or behaviors (SITB) to (1) determine the extent these methods have been applied and (2) summarize their findings. Through reviewing this body of literature, we also aimed to identify important future directions in the field and comment on how ML techniques may be used to improve suicide risk identification, in addition to clinical decision making and intervention.

3. Method

Our electronic search targeted papers published through February 2018 in the following databases: PsycINFO, PsycARTICLES, ERIC, CINAHL, and MEDLINE. Search terms included were: (a) “data mining” or “statistical learning” or “machine learning” or “big data” or “exploratory analyses”; (b) “self-harm” or “non-suicidal self-injury” or “nonsuicidal self-injury” or “NSSI” or “self-injury” or “self-injurious behaviors” or “self-mutilation” or “deliberate self-harm” or “cutting” or “self-cutting” or “self-burning” or “self-poisoning”; and (c) “suicide” or “self-injury” or “suicidality” or “self-harm” or “suicide” or “suicidal behavior” or “suicide attempt” or “suicide death” or “suicide plan” or “suicide thoughts” or “suicide ideation” or “suicide gesture” or “suicide threat.”

3.1. Inclusion criteria

The following inclusion criteria were employed: (a) inclusion of one or more of the following outcomes: non-suicidal self-injury, suicidal ideation, suicide planning, suicide attempt, suicide death; (b) employment of a machine learning technique to predict a SITB outcome; (c) inclusion of original empirical data; (d) written in English; and (e) peer reviewed (e.g., could not be a published dissertation study).

3.2. Data analyses implemented

In characterizing the types of ML analyses performed, we divided techniques into three over-arching categories or camps. The traditional delineation in ML is between unsupervised (e.g., clustering, principal components analysis) and supervised (e.g., regression, decision trees, random forests). However, since no papers in the current review used unsupervised learning, we only detail supervised methods. The first is the use of regularized regression (also known as penalized regression, lasso, elastic net, shrinkage and others). This family of methods is similar to the use of ordinary least squares but builds in a penalty term specifically to remove predictors from the model. This can be viewed as a more contemporary form of variable selection (as opposed to stepwise regression).

The second camp is decision trees, which can be characterized as an interpretable nonlinear method. A tree structure automatically incorporates interactions between predictors, along with step functions, to model nonlinear effects. In the creation of a decision tree, a subset of the predictors (those that are related to the outcome) are used to split on to maximize the variance explanation (or reduce misclassification with a categorical outcome) of the dependent variable, which in turn creates splits at optimal cutoff values of predictors. After a tree is constructed, with splitting occurring until misfit can no longer be reduced, it is common to “prune” the tree structure, creating smaller trees that can demonstrate increased generalizability while being easier to interpret. Tree methods that have a singular outcome create predictions that allow for comparisons with the actual outcome of interest, allowing for the calculation of accuracy in the case of categorical outcomes, or r-squared in the case of continuous outcomes. More complex models, such as structural equation model (SEM) trees, keep the tree structure, but splits are based on the SEM as the outcome. Similar to the use of mixture models, this allows for more complex group differences, such as differing slopes in a latent growth curve model.

In the third camp fall less interpretable (in some cases “black box”) methods that have increased predictive power over both penalized regression and decision trees. For example, random forests and boosting fit an ensemble (e.g., hundreds or thousands) of decision trees to increase the predictive power, while losing the interpretable tree structure. Although the interpretable tree structure is sacrificed, both boosting and random forests produce variable importance measures, a ranking of which predictors are used more and improve prediction across the hundreds (or thousands) of trees. Support vector machines increase the dimensionality of the model by creating new combinations of predictors in an attempt to find an optimal line (hyperplane) that can better bisect the classes.[2] With each of the methods that fall into the third camp, some degree of inference is sacrificed in favor of prediction.

Some articles in the current review used multiple methods falling in more than one camp, which allowed for comparing the use of more interpretable methods (e.g., elastic net regression, decision trees) to what can be gained by less interpretable methods (e.g., random forests, support vector machines). Given that several included articles used multiple methods, the employed methods are outlined in Table 1, as opposed to explicitly stated throughout the results section.

4. Results

4.1. Data extraction

Of the 288 studies produced by the search, 26 articles met inclusion criteria. The reference sections of these articles were examined, as were relevant articles in the literature, for additional, potentially pertinent articles not included in the initial electronic search (see Fig. 1). The final sample (35 articles) was further evaluated for type of outcome and grouped by category (articles could be included in more than one category). Studies were considered to fall into the suicide risk category if there was not a delineation between SITBs. See Table 1 for details of identified studies, including outcome category, sample description, study design, analyses used, and indicator information (when

[2] This is related to Fisher's linear discriminant analysis (LDA), where LDA also creates a hyperplane to separate classes, albeit with much more restrictive assumptions.

Table 1
Overview of included studies.

Authors, Year	Sample Description	Sample Size (Overall N; Target N)	Study Design	Outcome Variable	Statistical Analysis	Number of Indicators	Indicator Categories	Model Statistics
Suicide death								
Ilgen et al., 2009	Veterans diagnosed with depression[3,6] (U.S.)	887,859; 1,892	Longitudinal (63 months)	Suicide death	Bayesian dirichlet equivalent decisions tree (CV: training dataset)	8	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Treatment history,	+
Kessler et al., 2017b	Active duty service members with mental health visit and diagnosis of mental disorder or life difficulties[4,6] (U.S.)	33,880 (40,827 visits); 569	Longitudinal (5–26 weeks)	Suicide death	Naïve Bayes, Random forests, Support vector regression, Elastic net penalized regression (CV: training dataset, internal independent test dataset)	753–999	Demographics, Externalizing psychopathology, General psychopathology, Internalizing psychopathology, Military characteristics, Physical health, Prior SITBs, Psychosis, Social factors, Treatment history	AUC = 0.72
Kessler et al., 2017a	Veterans[3,6] (U.S.)	2,108,496; 6,359	Longitudinal (730 days)	Suicide death	Elastic net, Decision tree (Bayesian additive regression trees, Random forest), Spline (Adaptive splines, Adaptive polynomial splines), Generalized boosted regression models (Adaptive boosting), Support vector machines (Linear kernel, Polynomial kernel, Radial kernel) (CV: training dataset, internal independent test dataset)	381	Demographics, General psychopathology, Externalizing psychopathology, Internalizing psychopathology, Military characteristics, Physical health, Prior SITBs, Social factors, Treatment history	Sensitivity = 0.11
Kessler et al., 2015	Active duty service members with baseline psychiatric hospitalization[1,6] (U.S.)	40,820; 68	Longitudinal (12 months)	Suicide death	Regression trees, Elastic net regression, Least absolute shrinkage and selection operator (CV: 10-fold)	421	Cognition, Demographics, General psychopathology, Externalizing psychopathology, Internalizing psychopathology, Military characteristics, Prior SITBs, Psychosis, Social factors, Treatment history	AUC = 0.89
Poulin et al., 2014	Veterans[3,6] (U.S.)	210; 70	Longitudinal (12 months)	Suicide death	Supervised training with genetic programming; unspecified machine learning algorithm (unspecified) (CV: k-fold; training dataset)	+	Linguistic features	Accuracy = 69%
Suicide AttemptPast								
Baca-Garcia et al., 2010	Patients from psychiatric emergency room or hospitalization[1, 2] (U.S.)	277; 126	Cross-sectional	Lifetime suicide attempt	Forward selection, Support vector machine (CV: bootstrapping, training dataset)	840	Biology	Accuracy = 69%, sensitivity = 0.54, specificity = 0.80, positive likelihood ratio = 2.71, negative likelihood ratio = 1.75
Bae et al., 2015	Middle and high school students[5,6] (Korea)	2,754; +	Cross-sectional	Past year suicide attempt	Decision tree (CV: training dataset)	21	Demographics, Externalizing psychopathology, Internalizing psychopathology, Normative personality traits, Prior SITBs, Social factors	Accuracy = 90%
Burke et al., 2018*	Undergraduate students with history of NSSI[5] (U.S.)	359; 51	Cross-sectional	Lifetime suicide attempt	Elastic net regression, Decision tree, Random forests (CV: 10-fold, training dataset)	64	Demographics, Internalizing psychopathology, Prior SITBs	AUC = 0.75

(*continued on next page*)

Table 1 (*continued*)

Authors, Year	Sample Description	Sample Size (Overall N; Target N)	Study Design	Outcome Variable	Statistical Analysis	Number of Indicators	Indicator Categories	Model Statistics
Delgado-Gomez et al., 2016	Mental health inpatients and outpatients[1,3] (Spain)	902; 356	Cross-sectional	Recent suicide attempt	Decision tree (CV: 25 CVs, training dataset)	26	Demographics, Externalizing psychopathology, Internalizing psychopathology, Normative personality traits, Physical health, Prior SITBs, Social factors	Accuracy = 81.4%, sensitivity = 0.87, specificity = 0.86, precision = 0.86
Hettige et al., 2017	Patients with schizophrenia spectrum disorder[3] (Canada)	345; 131	Cross-sectional	Lifetime suicide attempt	Least absolute shrinkage and selection operator, Random forest, Support vector classifier, Elastic net (CV: stratified 10-fold, training dataset)	27	Demographics, General psychopathology, Externalizing psychopathology, Family history of psychopathology or SITBs, Normative personality traits, Psychosis, Social factors, Treatment history	AUC = 0.71, accuracy = 0.67, sensitivity = 0.64, specificity = 0.68
Just et al., 2017*	Undergraduate students with current suicidal ideation[5] (U.S.)	17; 9	Cross-sectional	Lifetime suicide attempt	Multivoxel analysis (CV: 1,000 random selections, training dataset)	+	Biology	Accuracy = 94%, sensitivity = 1.0, specificity = 0.88, PPV = 0.90, NPV = 1.0
Kuroki, 2015*	Filipino American community sample with lifetime suicidal ideation[4,6] (U.S.)	87; 34	Cross-sectional	Lifetime suicide attempt	Random forest (CV: not fully specified)	31	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Social factors	+
Kuroki and Tilley, 2012	Asian American community sample with lifetime suicidal ideation[4,6] (U.S.)	191; 56	Cross-sectional	Lifetime suicide attempt	Decision tree (CV: training dataset)	30	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Social factors	Sensitivity = 0.75, specificity = 0.39, PPV = 0.39, NPV = 0.75
Lopez-Castroman et al., 2011	Patients presenting with a SA at Emergency Department[2] (Spain, France)	1,349; 617	Cross-sectional	Three or more lifetime suicide attempts	Incremental association markov blanket; support vector machine; Max-min hill-climbing (CV: 10-fold, training dataset)	46	Demographics, Externalizing psychopathology, Family history of psychopathology or SITBs, General psychopathology, Internalizing psychopathology, Prior SITBs, Psychosis	+
Mann et al., 2008	Patients with mood, schizophrenia spectrum, or personality disorders[1,3] (U.S.)	290; 80	Cross-sectional	Recent suicide attempt, Remote suicide attempt	Decision tree (CV: 10-fold, training dataset)	21	Demographics, General psychopathology, Internalizing Psychopathology, Externalizing psychopathology, Family history of psychopathology or SITBs, Physical health, Normative personality traits, Social factors, Prior SITBs, Psychosis	AUC = 0.80, sensitivity = 0.73, specificity = 0.80, PPV = 0.58
Metzger et al., 2017	Patients from Emergency Department[2,6] (France)	390; 118 921; 307	Cross-sectional	Lifetime suicide attempt	Random forest, Naive bayes, Support vector machines, Predictive association rules, Decision trees, Neural networks (CV: training dataset)	+	Demographics, Externalizing psychopathology, General psychopathology, Internalizing psychopathology, Linguistic features, Physical health	Sensitivity = 0.95, PPV = 0.97
Oh et al., 2017	Patients from outpatient mental health[5]	573; 163	Cross-sectional	Lifetime suicide attempt, Past year suicide attempt, Past month suicide attempt	Artificial neural network (CV: training dataset)	41	Demographics, General psychopathology, Internalizing Psychopathology, Externalizing psychopathology, Physical health, Normative personality traits, Social factors, Prior SITBs	AUC = 0.93, accuracy = 93.7%, sensitivity = 0.12, specificity = 0.99

(*continued on next page*)

Table 1 (*continued*)

Authors, Year	Sample Description	Sample Size (Overall N; Target N)	Study Design	Outcome Variable	Statistical Analysis	Number of Indicators	Indicator Categories	Model Statistics
Passos, Mwangi, Cao, Hamilton, Wu, Zhang, … & Soares, 2016	Patients with major depressive disorder or bipolar disorder[3] (U.S.)	144; 43	Cross-sectional	Lifetime suicide attempt	Least absolute shrinkage and selection operator, Support vector machines, Relevance vector machine (CV: 10-fold, leave-one-out, training dataset)	16	Demographics, Externalizing psychopathology, General psychopathology, Internalizing psychopathology, Treatment history	AUC = 0.77, sensitivity = 0.72, specificity = 0.71
Walsh et al., 2017	Patients at community hospital[3,6] (U.S.)	5,167; 3,250	Longitudinal (1 week - 24 months)	Prospective suicide attempts	Random forests (CV: bootstrapping, training dataset)	1328	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Prior SITBs, Treatment history	AUCs = 0.80–0.84
Suicide Planning								
Burke et al., 2018*	Undergraduate students with a history of non-suicidal self-injury[5] (U.S.)	359; 33	Cross-sectional	Past week suicide plan	Elastic net regression, decision tree, random forests (CV: 10-fold, training dataset)	62	Demographics, Internalizing psychopathology, Prior SITBs	AUC = 0.89
Suicidal Ideation								
Batterham and Christensen, 2012	Adult community sample[4,6] (Australia)	6656; 406	Longitudinal (48 months)	Past year suicidal ideation	Decision tree (CV: not fully specified)	35	Demographics, Internalizing Psychopathology, Externalizing psychopathology, Physical health, Normative personality traits, Social factors, Physical health, Prior SITBs	+
Burke et al., 2018*	Undergraduate students with a history of non-suicidal self-injury[5] (U.S.)	359; 90	Cross-sectional	Past week suicidal ideation	Elastic net regression, decision tree, random forests (CV: 10-fold, training dataset)	62	Demographics, Internalizing psychopathology, Prior SITBs	AUC = 0.85
Cook et al., 2016	Patients from Emergency Department or hospitalization [1,2] (Spain)	1,453; 844	Longitudinal (12 months)	Suicidal ideation during study period	Natural language based processing based machine learning (CV: training dataset)	+	Internalizing psychopathology, Linguistic features, Treatment history	Sensitivity = 0.76, specificity = 0.62, PPV = 0.73
Gradus et al., 2017	Veterans[3] (U.S.)	2,088; 370	Cross-sectional	Suicidal ideation since most recent deployment	Decision tree, Random forests (CV: bootstrapping)	25	Demographics, Externalizing psychopathology, General psychopathology, Internalizing psychopathology, Military characteristics, Social factors	AUC = 0.92
Handley et al., 2014	Older adult community sample[4,6] (Australia)	2160; 95	Longitudinal (60 months)	Past 2-week suicidal ideation	Decision tree (CV: 10-fold, K-fold)	21	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Social factors	AUC = 0.81
Hill et al., 2017	Adolescent community sample[4,6] (U.S.)	4,799; 523	Longitudinal (12 months)	Past 12-month suicidal ideation	Decision tree (CV: 10-fold)	23	Cognition, Demographics, Exposure to SITBs, Family history of psychopathology or SITBs, Internalizing psychopathology, Prior SITBs, Social factors	Accuracy = 80.10%, sensitivity = 0.63, specificity = 0.82
Jordan et al., 2018	Patients from primary care[3] (U.S.)	6,805; 857	Cross-sectional	Past two-week suicidal ideation	Decision tree, Support Vector Machine, Neural network, Fisher's linear discriminant function approach (CV: training dataset)	4; 31; 3	Demographics, Screeners	AUC = 85.6, sensitivity = 0.78, specificity = 0.83, PPV = 0.39, NPV = 0.97
Just et al., 2017*	Undergraduate students[5] (U.S.)	34; 17	Cross-sectional	Current suicidal ideation	Multivoxel analysis (CV: 1,000 random selections, training dataset)	+	Biology	Accuracy = 91%, sensitivity = 0.88, specificity = 0.94, PPV = 0.94, NPV = 0.89

(*continued on next page*)

Table 1 (*continued*)

Authors, Year	Sample Description	Sample Size (Overall N; Target N)	Study Design	Outcome Variable	Statistical Analysis	Number of Indicators	Indicator Categories	Model Statistics
Kuroki, 2015*	Filipino American community sample[4, 6] (U.S.)	624; 87	Cross-sectional	Lifetime suicidal ideation	Random forest (CV: not fully specified)	30	Demographics, Externalizing psychopathology, Internalizing psychopathology, Prior SITBs, Physical health, Social factors	+
Kuroki and Tilley, 2012	Asian American community sample[4,6] (U.S.)	2095; 191	Cross-sectional	Lifetime suicidal ideation	Balanced random forest (CV: bootstrapping, training dataset)	30	Demographics, Externalizing psychopathology, Internalizing psychopathology, Physical health, Social factors	Sensitivity = 0.72, specificity = 0.76, PPV = 0.23, NPV = 0.96
Suicide Risk[a]								
Barros et al., 2017	Patients with mood disorder from outpatient and inpatient mental health[3] (Chile)	707; 349	Cross-sectional	Sought treatment for suicide attempts or current suicidal ideation	AdaBoost, Decision tree, K-nearest neighbor, Random forest, Neural-network multilayer perceptron, Support vector machine (CV: 10-fold)	343	Demographics, Internalizing psychopathology, Social factors	Accuracy = 78%, sensitivity = 0.77, specificity = 0.78
Batterham and Christensen, 2012	Adult community sample[4,6] (Australia)	6656; 99	Longitudinal (48 months)	Past year suicide plans or suicide attempt	Decision tree (CV: not fully specified)	35	Demographics, Internalizing Psychopathology, Externalizing psychopathology, Physical health, Normative personality traits, Social factors, Physical health, Prior SITBs	+
Braithwaite et al., 2016	Community sample of social media users[4] (U.S.)	135; 17	Cross-sectional	Depressive Symptom Inventory – Suicide Subscale total score > 2	Decision tree (CV: leave-one-out; training dataset)	135	Linguistic features	Sensitivity = 0.53; specificity = 0.97; PPV = 0.75; NPV = 0.93
Cheng et al., 2017	Community sample of social media users[4] (China)	974; 197	Cross-sectional	Suicide Probability Scale total score ≥ 80 or told others via social media in the past 12 months that he/she wanted to kill self	Support vector machine (CV: leave-one-out; training dataset)	72	Linguistic features	AUC = 0.61, sensitivity = 0.65, specificity = 0.58
Guan et al., 2015	Community sample of social media users[4] (China)	909; 144	Cross-sectional	Suicide Probability Score total score > 1 standard deviation above mean	Random forest (CV: 5-fold; training dataset)	88	Demographics, Linguistic features; Social media use features	Recall value = 0.82
Morales et al., 2017	Patients receiving mental health treatment[1,3] (Chile)	707; 349	Cross-sectional	Consultations relating to suicide attempt or presenting suicidal ideation in preceding years	Cross industry standard process for data mining, Decision tree (CV: training dataset)	345	Demographics, Externalizing psychopathology, Internalizing psychopathology, Social factors	AUC = 0.59, accuracy = 0.71, precision = 0.73, recall value = 0.63, specificity = 0.79
Pestian et al., 2016	Adolescents from Emergency Department[2] (U.S.)	60; 30	Cross-sectional	Presented to emergency department with suicidal ideation, gestures, or attempts	Cosin support vector machine (CV: leave-one-out)	+	Linguistic features	Accuracy = 96.67%
Pestian, Sorter, Connolly, Cohen, McCullumsmith, Gee, … & Rohlfs, 2017	Adolescents and adults from Emergency Department and inpatient and outpatient mental health[1,2,3] (U.S.)	379; 130	Cross-sectional	Presented to emergency department or inpatient unit with suicidal ideation or attempts within previous 24 hours	Support vector machine (CV: leave-one-out)	+	Linguistic features	AUC = 0.93
Non-Suicidal Self-Injury								
Ammerman et al., 2017a	Undergraduate students[5] (U.S.)	3,559; 428	Cross-sectional	Frequency of non-suicidal self-injury	Structural equation modeling trees (CV: not mentioned)	7	Internalizing psychopathology, Prior SITBS	+
Ammerman et al., 2018	Undergraduate students with a history of non-suicidal self-injury[5] (U.S.)	957; 957	Cross-sectional	Non-suicidal self-injury age of onset	Decision trees (CV: not mentioned)	7	Prior SITBs	+

(*continued on next page*)

Table 1 (*continued*)

Authors, Year	Sample Description	Sample Size (Overall N; Target N)	Study Design	Outcome Variable	Statistical Analysis	Number of Indicators	Indicator Categories	Model Statistics
Ammerman et al., 2017b	Undergraduate students with a history of non-suicidal self-injury[5] (U.S.)	712; 712	Cross-sectional	Frequency of non-suicidal self-injury	Lasso regression; Random forests (CV: bootstrapping)	26	Externalizing psychopathology, Internalizing psychopathology, Prior SITBs	$R^2 = 0.48$

Note: SITB = Self-injurious thoughts and behaviors;
* studies were included in more than one category due to having more than one SITB outcome;
+ = not reported / unable to determine; Superscripts in Sample Description represent the setting from which participants were recruited and/or data was collected:
1 = Psychiatric inpatient hospital,
2 = Emergency department,
3 = Outpatient setting (e.g., psychiatric outpatient clinic, general hospital),
4 = Community setting (e.g., general community recruitment),
5 = Academic setting (e.g., high school, university),
6 = Data from existing database (e.g., national database, electronic medical record); Country of data collection is included in parentheses in Sample Description column;
a = definition of suicide risk is reported as defined by the authors; CV = cross-validation; Decision trees may also be referred to as classification trees, regression trees, and recursive partitioning

available). Indicators (predictors) were classified into 18 different broad categories (adapted from Franklin et al., 2017) (see Table 2).

4.2. Performance metric(s) extraction

Across the different types of methods, the performance metrics that should be reported depend on the distribution of the outcome. For single, categorical outcomes, metrics such as accuracy and the area under the receiver operating characteristic curve (AUC) capture how well the classes are predicted. Whereas accuracy needs to be examined with respect to the class proportion, the AUC marries both sensitivity (prediction of the positive class) and specificity (prediction of the negative class). AUC can be examined graphically, as the curve depicts the sensitivity and specificity values across the range of cutoffs for class membership. Most machine learning methods create predicted class probabilities, then requiring the user to either use the default probability of 0.5 (those with predicted probabilities > 0.5 are assigned to class one) or use a different cutoff that better reflects the cost for misclassifying both classes. For both accuracy and AUC, values closer to one indicate better performance, however, more specific metrics should be examined (and reported) to determine how this relates to predicting each class or where the model is "off". Additionally, there are more fine-grained metrics, such as Positive Predictive Values (PPV) or Negative Predictive Values (NPV), that measure the accuracy of the predicted values of the positive and negative conditions, as opposed to the true condition (as in sensitivity and specificity). For imbalanced outcomes, it is often more informative to focus on metrics such as PPV (also known as precision) or sensitivity (also known as recall) that quantify the quality of prediction for only the positive cases. For single, continuous outcomes, metrics such as the root mean squared error (RMSE) or r-squared are generally reported. Methods that model more complex forms of outcomes can result in improvements in the log-likelihood (as in having a structural equation or multilevel model), or other performance metrics.

An additional component of reporting performance metrics is whether they are calculated on the entire sample, a holdout (i.e., test) sample, or through using cross-validation (CV) or bootstrapping, or a combination. Although the evaluation performance on a holdout dataset (treating the model as fixed "predicting" on the never before used sample), there are drawbacks to this approach when the sample size is not large (e.g. Steyerberg and Harrell, 2016). The use of CV and/or bootstrapping is generally used for two purposes: to select the optimal values of the tuning parameters, and to derive a more realistic estimate of model performance. Most studies reported using some form of CV to assess model performance, however, less information was reported for the selection of tuning parameter. Additionally, when pairing CV or bootstrapping with an imbalanced outcome, it may be necessary to *stratify* the assignment of the minority class to be evenly distributed across the folds (or bootstrap samples). Few studies reported the use of stratified CV.

Even studies that focus mostly on inference should, in most cases, provide model performance metrics, as model interpretations should be examined in light of the prediction performance. Not all included articles reported a model performance metric; however, when available, they are reported in the results section. If multiple model performance metrics were reported (i.e., for each of the multiple models), only the metrics for the best performing models and/or the model highlighted by the study authors are reported (see Table 1). Given the inconsistency of model performance statistics across ML techniques, it was not appropriate to conduct a meta-analysis; instead, main findings from the included articles are outlined below.

4.3. Suicide death

Our systematic review identified five studies that used ML techniques to predict suicide death. All studies utilized either U.S. veteran or

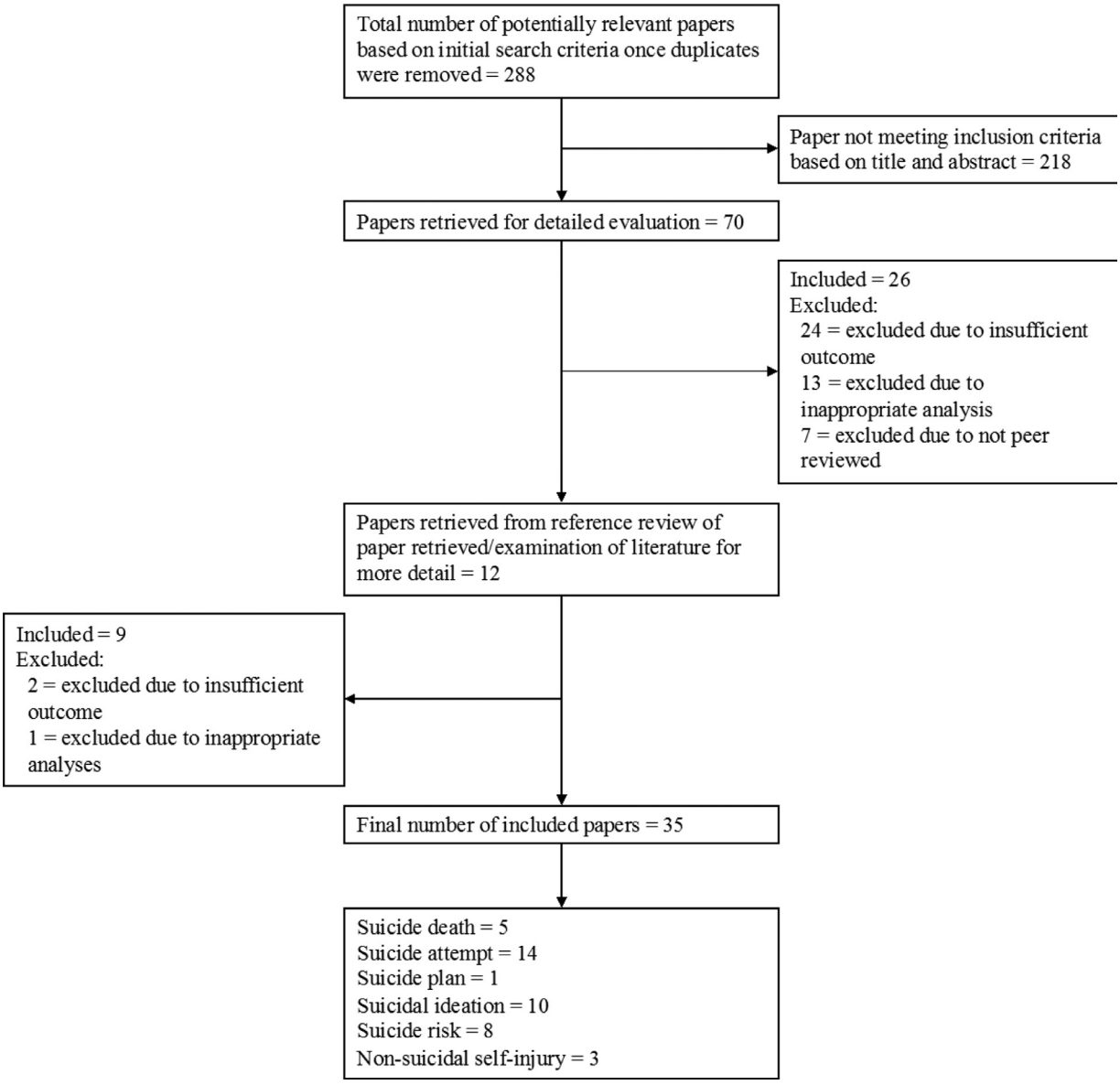

Fig. 1. Study inclusion flow chart.

Table 2
Classification categories for indicators.

1. Biology (e.g., genes)
2. Screeners (e.g., screening instruments, patient prediction)
3. Cognition (e.g., problem-solving, intelligence)
4. Demographics (e.g., gender, age)
5. Externalizing psychopathology (e.g., substance abuse, aggression)
6. Family history of psychopathology (e.g., relative suicide attempt, maternal depression)
7. General psychopathology (e.g., number of psychiatric diagnoses)
8. Internalizing psychopathology (e.g., mood disorders, hopelessness)
9. Linguistic features (e.g., language characteristics of social media posts or clinical notes, linguistic and acoustic responses to open-ended questions)
10. Military characteristics (e.g., number of deployments)
11. Normative personality traits (e.g., openness)
12. Physical health (e.g., migraines, weight)
13. Psychosis (e.g., schizophrenia)
14. Prior SITBs (e.g., presence and features of previous SITBs)
15. Exposure to SITBs (e.g., friend suicide attempt)
16. Social factors (e.g., stressful life events, abuse history)
17. Social media use features (e.g., number of posts)
18. Treatment history (e.g., hospitalizations, specific medication use)

service member samples and employed a longitudinal design. The majority of studies used electronic medical record (EMR) variables as indicators and included between 8 and 979 (M = 447.25) indicators in their models. Most of the identified studies employed multiple ML techniques to predict suicide death (M = 3.6; Range = 1–9).

First, three studies utilized ML to improve prediction of suicide death among those with varying psychiatric histories (Kessler et al., 2015; Kessler et al., 2017a, 2017b). Among service members with baseline psychiatric hospitalization, ML model AUCs ranged from 0.71–0.89 in predicting suicide death (Kessler et al., 2015). Findings demonstrated that certain demographic (e.g., male, enlisting at 27 or older), contextual (e.g., having access to firearms), and psychiatric and behavioral history (e.g., crime perpetration, SITB history, prior psychiatric treatment) factors were selected as important in the model. Similarly, among service members with a baseline outpatient mental health visit, ML models predicting suicide 26 weeks after index outpatient visit demonstrated AUCs of 0.72 for service members with prior psychiatric hospitalization(s), 0.61 for service members without prior hospitalizations, and 0.66 for a combined sample (Kessler et al., 2017b). AUCs improved when predicting suicide deaths five weeks after index visit. History of SITBs and mental health treatment characteristics

were identified as important indicators among service members with prior psychiatric hospitalization, whereas physical health conditions and recent crime perpetration emerged as important indicators for those without psychiatric hospitalization, suggesting possible divergent causal processes for these subgroups (Kessler et al., 2017b). Finally, research was extended to veterans who died by suicide and time-matched controls. While authors did not discuss important indicators in their models, results demonstrated that the highest performing ML model evidenced a sensitivity of 2.7% and 10.7% among veterans identified at the top 0.1% and 1% of suicide risk within the healthcare system, respectively (Kessler et al., 2017a).

Research predicting suicide death has also utilized ML to identify risk factor interactions to classify those at high-risk for suicide (Ilgen et al., 2009). Among veterans diagnosed with depression, a three-way interaction was identified, where individuals who identified as non-African American with a substance use disorder and who were psychiatrically hospitalized within the past year were at highest risk for suicide (Ilgen et al., 2009). A final study among veterans utilized natural language processing to identify important words in unstructured clinical notes that distinguished those who died from suicide. Words highlighted as important fell under the categories of patient behavior (e.g., agitation, frightened), physical health conditions (e.g., cardiac, gastrointenstinal) and care descriptors (i.e., integrated) (Poulin et al., 2014). For models including only single words, accuracy estimates ranged from 46–65%; for models including phrases, accuracy estimates improved, ranging from 52–69% (Poulin et al., 2014).

4.4. Suicide attempt

Our systematic review identified 14 studies that used ML techniques to predict SAs. The adult sample characteristics varied, and one study examined an adolescent sample. All studies were cross-sectional in design, with the exception of one longitudinal study. The majority of studies utilized EMR variables as indicators. The identified studies included between 16 and 1328 (M = 207.58) indicators in their models. Approximately half of the studies employed multiple ML techniques to predict SA (M = 2.07; Range = 1–7).

The single identified longitudinal study utilized a sample of adults who had a EMR documented self-injury code (Walsh et al., 2017). Researchers developed ML models based on EMR indicators to predict SA over varying timeframes (AUCs = 0.80–0.84), with model performance increasing closer to time of SA (from 720 days to 7 days prior to the SA) (Walsh et al., 2017). Recurrent depression with psychosis, schizophrenia, and schizoaffective disorder, in addition to evidence of prior SITB, were consistently ranked as important. Age and substance use dependence increased in importance in shorter timeframes, whereas certain medication classes (e.g., serotonin reuptake inhibitors, benzodiazepines) appeared more important over greater timeframes (Walsh et al., 2017).

Four studies used samples of adults who were either hospitalized or admitted to the emergency department due to suicidal behavior and/or psychiatric concerns. Delgado-Gomez et al. (2016) employed ML to differentiate between adults presenting with a first SA and those with no SA. The accuracy of the model was 81.4% (sensitivity = 0.87; specificity = 0.86; precision = 0.86), and indicators identified as important in the model included feelings of emptiness, arguments with spouse, tantrums/angry outbursts, history of adult SITB, and self-control (Delgado-Gomez et al., 2016). Similarly, ML was used to classify presence of SA versus controls (which included participants reporting SI) (Metzger et al., 2017). The two best performing models achieved sensitivity of 0.95–0.96 and PPV of 0.93–0.97; important indicators in these models were not outlined (Metzger et al., 2017). Similar methods were also used to differentiate between single versus repeated SAs, however, findings focused on identifying the most influential indicators in the models (Lopez-Castroman et al., 2011). Authors highlighted age as an important indicator (sensitivity = 0.51; specificity = 0.97; precision = 0.76), in addition to the presence of an anxiety disorder, alcohol/drug use diagnosis, marital status, and previous SITB characteristics (average AUC = 0.71; Lopez-Castroman et al., 2011). Finally, Baca-Garcia et al., (2010) extended findings to examine biological factors that may predict lifetime SA among males. ML algorithms included 840 total single nucleotide polymorphisms selected from 312 central nervous system genes. Three single nucleotide polymorphisms were identified as having the most explanatory power in SA classification, which, when considered together, accurately classified 69% of the sample (sensitivity = 0.54, specificity = 0.80, positive likelihood ratio = 2.71, negative likelihood ratio = 1.75); these estimates were found to remain relatively stable in a replication sample (Baca-Garcia et al., 2010). An additional fifth study employed ML to predict recent and remote SA history among a sample diagnosed with mood, schizophrenia spectrum, or personality disorders (Mann et al., 2008). Mann et al. (2008) found that current SI was the most important predictor of recent SA status (AUC = 0.80, sensitivity = 0.73, specificity = 0.80, PPV = 0.58) whereas lifetime aggression was the most important predictor of remote SA status (AUC = 0.65, sensitivity = 0.89, specificity = 0.36, PPV = 0.44). Borderline personality disorder and depression further differentiated positive and negative cases among recent and remote attempters, respectively (Mann et al., 2008).

The three following studies included samples receiving outpatient mental health care. Passos et al. (2016) recruited adults with major depression or bipolar disorder diagnoses and classified lifetime SA history. Accuracy of the methods ranged from 65% to 72%, with the highest accuracy model evidencing an AUC of 0.77, 0.72 sensitivity, and 0.71 specificity. Across models, several indicators were identified as most important, including previous hospitalization(s) for depression, a history of psychosis, and comorbid conditions of cocaine dependence and post-traumatic stress disorder (Passos et al., 2016). Similar findings were demonstrated in the prediction of lifetime SA among a sample of adults diagnosed with schizophrenia spectrum disorders. Of all classifiers, the most predictive model observed the following performance metrics: AUC = 0.71, accuracy = 0.67, sensitivity = 0.64, specificity = 0.68. Important indicators in the models included duration of illness, number of hospitalizations, childhood emotional and physical abuse, and lifetime drug abuse/dependence (Hettige et al., 2017). Finally, the current review identified a study that classified past one-month, one-year, and lifetime SA among individuals with depression and anxiety disorders. Using 31 self-report psychiatric questionnaires, ML models were found to have greatest accuracy of predicting past one-month SAs (AUC = 0.93, accuracy = 93.7%, sensitivity = 0.12, specificity = 0.99; past one-year model: AUC = 0.89, accuracy = 90.8%, sensitivity = 0.33, specificity = 0.98; lifetime model: AUC = 0.87, accuracy = 87.4%, sensitivity = 0.77, specificity = 0.91). Of all scales included, the Emotion Regulation Questionnaire exhibited the greatest contribution to classification performance (Oh et al., 2017).

Although the five final studies predicting SA all utilized community samples, they were novel in sample subsets of interest or in collection methods. The first study was the only to focus on adolescents (i.e., middle and high school students; Bae et al., 2015). The use of ML techniques resulted in a model with 90% accuracy in predicting past year SA. Factors identified as important included depression, delinquency, family intimacy, and stress. Another study supported the role of interpersonal factors in predicting lifetime SA history among a community sample of Filipino American adults with a SI history. The two most important predictors identified were the number of family relatives living within a 90-minute drive and family conflict (Kuroki, 2015). A similar study examined a nationally representative sample of Asian Americans and found that among those with a history of lifetime SI, the important predictors of lifetime SA identified were family conflict, family support, and unfair treatment due to discrimination (sensitivity = 0.75, specificity = 0.39, PPV = 0.39, NPV = 0.75) (Kuroki and Tilley, 2012). Using primarily NSSI

characteristics as model indicators, another study employed ML techniques to predict lifetime SA among undergraduates with a history of NSSI. The model with greatest performance achieved an AUC of 0.75. Despite the inclusion of SI and SP as indicators, variables selected as important across ML models were NSSI characteristics (i.e., anti-suicide function of NSSI, number of NSSI scars, and history of medical treatment due to NSSI) (Burke et al., 2018). In an attempt to differentiate undergraduate students with SI from those with SA, functional magnetic resonance imaging (fMRI) was used to examine group differences in the neural representations of positive, negative, and suicide-related concepts. Using ML, authors identified a model that differentiated those reporting a SA with 94% accuracy (sensitivity = 1, specificity = 0.88, PPV = 0.90, NPV = 1). The concepts that best discriminated between groups were 'death', 'lifeless', and 'carefree' (Just et al., 2017).

4.5. Suicide planning

Our systematic review identified one study that used ML techniques to predict recent SP among those with a history of NSSI, utilizing 62 indicators, primarily reflecting NSSI characteristics. The model with the greatest performance achieved an AUC of 0.89. Across the ML models, results suggested that depressive symptoms and the endorsement of the anti-suicide function of NSSI were the two most important predictors (Burke et al., 2018).

4.6. Suicidal ideation

Our systematic review identified ten studies that used ML techniques to predict SI. Studies utilized various samples, ranging from adult primary care patients to community adult and adolescent samples. The majority of studies employed cross-sectional designs, however four studies utilized longitudinal designs. Models incorporated between 3 and 62 indicators (M = 32.13) and utilized one to four ML methods (M = 1.6).

In a longitudinal framework among an adolescent sample, Hill et al. (2017) used ML to predict SI at a one-year follow-up. Authors highlighted three different, well-performing models (sensitivity range = 0.47–0.78; specificity range = 0.68–0.91). The model with moderate sensitivity and high specificity included indicators reflecting depressive symptoms, family/peer suicide, and social support, whereas the most sensitive, but least specific, model used these same factors, in addition to gender, ethnicity, hours of sleep, school-related factors, and future orientation (Hill et al., 2017). An additional longitudinal study employed natural language processing in a sample of adults after hospital discharge for a suicide-related event (Cook et al., 2016). Eight structured indicators, in addition to answers to an unstructured question (e.g., "how do you feel today?"), were included in their model to predict the presence of SI over follow-up ranging from two days to 12 months. Results suggested that from the unstructured question, the phrase "I told" was most important in predicting SI, followed by "monotony", "Ritalin", "harassed", and "we work"; the structured indicator most predictive of SI was older age, whereas rarely being angry and reporting greater wellbeing were associated with lower odds of SI. Results also showed that models including the structured indicators (sensitivity = 0.76, specificity = 0.62, PPV = 0.73), as opposed to responses from the unstructured questions (sensitivity = 0.56, specificity = 0.57, PPV = 0.61), performed slightly better in predicting SI (Cook et al., 2016). A third study used ML to predict recent SI among a sample of older adults followed for a period of five-years (Handley et al., 2014; AUC = 0.81). Authors found that psychological distress was the strongest predictor of SI at follow-up. Results indicated that whereas among those with high psychological distress, physical functioning emerged as an important indicator of SI, among those with low psychological distress, social support emerged as an important indicator of SI (Handley et al., 2014). A fourth study used ML to predict past year SI among a community sample of adults. Baseline SI emerged as the strongest predictor of SI at four-year follow-up (Batterham and Christensen, 2012). Among those with baseline SI, neuroticism differentiated those at lower and higher risk of SI at follow-up. Among those without baseline SI, anxiety emerged as an important indicator, with the tree structure suggesting that those experiencing the highest level of anxiety exhibit a threefold increase in SI risk (Batterham and Christensen, 2012).

Three cross-sectional studies identified in the review used large adult samples to predict SI. Among primary care patients, past two-week SI was predicted based on responses to three common self-report screeners used in primary care settings, in addition to sociodemographic variables. All ML models exhibited AUCs greater than 0.80, and, across models, sociodemographic variables added no value beyond the scales. Authors highlighted the simplest classification model among the best classifiers (overall AUC = 85.6: sensitivity = 0.78, specificity = 0.83, PPV = 0.39, and NPV = 0.97), which relied on four individual scale items assessing depressed mood, feelings of worthlessness, sleep problems, and uncontrollable worry (Jordan et al., 2018). Gradus et al. (2018) used ML techniques to examine potential gender differences in SI prediction among a large sample of veterans. Models observed AUCs of 0.91 and 0.92 for males and females, respectively; in both models, probable depression, post-traumatic stress, and anxiety disorder diagnoses, in addition to alcohol use, were particularly important variables. Further, among females, sexual harassment during deployment differentiated between those at higher and lower SI probabilities. Kuroki (2015) found that in employing ML to classify SI history among a community sample of Filipino American adults, depressive disorder, years in the United States, substance use disorder, and number of negative life events emerged as the most important predictors. Similarly, Kuroki and Tilley (2012) examined a nationally representative sample of Asian Americans and found that the most important predictors of lifetime SI were depressive and anxiety disorders, followed by family conflict and family cohesion (sensitivity = 0.72%, specificity = 0.76, PPV = 0.23, NPV = 0.96).

The final two studies predicting SI have been featured in the previous section(s) (one predicting both SA and SP, and one predicting SA). The first found similar results in predicting SI as it did in predicting SP among undergraduate students with a NSSI history. The model with greatest performance observed an AUC of 0.85, and, across ML techniques, results converged in identifying the anti-suicide function of NSSI and depression as important model indicators (Burke et al., 2018). Also among undergraduate students, fMRI was used to classify individuals with current SI, versus those without SI, based on neural representations of positive, negative, and suicide-related concepts. Results of the study produced a model that classified current SI with 91% accuracy (sensitivity = 0.88, specificity = 0.94, PPV = 0.94, NPV = 0.89). There was some overlap in important concepts discriminating between SI versus no SI compared to SA versus SI outcomes; the concepts of 'death', 'cruelty', 'trouble', 'carefree', 'good', and 'praise' were identified as important, with 'death' emerging as the most discriminating (Just et al., 2017).

4.7. Suicide risk

Our systematic review identified a total of eight studies that used ML techniques to predict suicide risk. The studies used community and clinical samples of adults, as well as clinical samples of adolescents. The majority of studies utilized linguistic features of social media posts and responses to unstructured questions. The studies included between 35 and 345 indicators in their models (M = 169.65). The majority of studies employed a singular ML technique to predict suicide risk (M = 1.63; Range = 1–6).

Barros et al. (2017) recruited mental health patients diagnosed with mood disorders to classify individuals considered to be at suicide risk. ML models largely incorporated sociodemographic and clinical data (patient-reported and EMR). The best performing model evidenced an

accuracy of 78% (sensitivity = 0.77; specificity = 0.78). The resultant models identified reasons for living, experience of unrest, and personal satisfaction as important variables in distinguishing between groups.

Three studies utilized both passive and active data collection methods for prediction, leveraging the use of social media. The first targeted information from social media post content (e.g., language processes, psychological processes) to classify participants considered at high suicide risk (Guan et al., 2015). Results demonstrated similar performance across two models, with the optimal model achieving a recall value (number of true positives/total number of positive instances) of 0.82 (Guan et al., 2015). A similar study utilized the social media platform Twitter to predict suicide risk (Braithwaite et al., 2016). Similarly, natural language processing was used to extract social media post content, which was included in the ML analysis. The model performance metrics were: sensitivity = 0.53; specificity = 0.97; PPV = 0.75; and NPV = 0.93. The results suggested that individuals posting fewer words related to "achieve", "religion" and "relativity" were more likely to be deemed at risk for suicide (Braithwaite et al., 2016). A third study used linguistic features of social media posts among Chinese adults to predict suicide risk (Cheng et al., 2017). The researchers found poor ML performance in predicting suicide risk and suicide-related posts. However, among the subset of users who had told others via social media that they wanted to kill themselves in the past 12 months, the ML model's performance improved (AUC = 0.61; sensitivity = 0.65; specificity = 0.58).

The four remaining studies in this category mainly used patient reported data in risk prediction. The first study predicted past year SP or SA (SP/SA) among a sample of adults followed for four-years. Baseline SP/SA emerged as the strongest predictor of follow-up SP/SA; no other variables emerged as important in distinguishing those with SP/SA at follow-up among those with baseline SP/SA (Batterham and Christensen, 2012). However, among those without baseline SP/SA, depression emerged as important in distinguishing risk for follow-up SP/SA, with neuroticism and marijuana use further distinguishing risk groups (Batterham and Christensen, 2012). The second study collected questionnaires of psychological constructs among outpatient mental health patients to predict suicide risk. The best performing model demonstrated the following metrics: AUC = 0.59; accuracy = 0.71; precision = 0.73; recall value = 0.63; specificity = 0.79. Incorporating all model findings, results identified that reporting thoughts of ending one's life, more frequent headaches, being frightened when feeling alone, greater dissatisfaction with life, feeling empty inside, and less fear of the act of killing oneself were important variables in group prediction (Morales et al., 2017). Using similar methodologies, the next two studies analyzed the linguistic responses to open-ended questions (e.g., "Does it hurt emotionally?", "Do you have hope?") and associated vocal characteristics to classify the likelihood of presenting to the emergency room for SITBs (i.e., SI or SA). Considering the analysis of just the linguistic responses to the open-ended questions, the best fitting model accurately classified 96.67% of adolescents who presented with SITB (Pestian et al., 2016). The final study of this nature used similar methodology among a combined sample of adolescents and adults. Authors incorporated linguistic responses and vocal characteristics of responses to the unstructured questions in ML models to differentiate between emergency room patients with SITBs, a psychiatric control group, and a non-psychiatric control group. The two best fitting models (AUC = 0.93) were those distinguishing between the SITB and control groups. The first model was among the combined adolescent and adult sample and used only linguistic responses, whereas the second model was among the adult sample using both linguistic and vocal responses. Utilizing both linguistic and vocal responses was most additive in the adolescent and adult combined sample when distinguishing between the SITB and psychiatric control participants (AUC = 0.82; Pestian et al., 2017).

4.8. Non-suicidal self-injury

Our systematic review identified three studies that used ML techniques to predict NSSI. All studies utilized an undergraduate student sample and included between 1 and 27 indicators in their models (M = 11.67), and, on average, used one ML model (M = 1.33, *Range* = 1–7).

Two articles in this section aimed to identify subgroups of individuals who engaged in NSSI by identifying splits, or numerical cut points, on NSSI-related variables. The first utilized a ML technique to examine splits in number of NSSI acts during the previous year as predicted by participant-reported psychological difficulties. Results demonstrated significant splits between zero and one past year NSSI acts (i.e., resulting in a subgroup that had not engaged in NSSI in the past year and those with one or more past year NSSI acts) and between five and six past year NSSI acts (i.e., resulting in a subgroup that reported one to five past year NSSI acts and a subgroup that reported six or more past year NSSI acts), suggesting that participants reporting six or more past year NSSI acts may represent a more severe group of self-injurers (Ammerman et al., 2017a). Another study utilized a similar approach to examine splits in NSSI behavior age of onset predicted by prior SITBs, including NSSI characteristics (e.g., NSSI frequency, number of NSSI-related hospital visits), SI, SP, and SA. Multiple ML models were used and, taken together, results suggested there is a potential subgroup in the data representing those with an earlier age of onset (i.e., approximately 12 or younger); this subgroup reported greater NSSI frequency, number of NSSI methods, and NSSI-related hospital visits, in addition to increased likelihood of having a SP (Ammerman et al., 2018). The final study employed two ML techniques to identify important indicators of NSSI frequency, both explaining a significant proportion of variance in NSSI frequency (R^2 = 0.48 and 0.46, respectively). Models indicated that the number of NSSI methods was the most important indicator of lifetime NSSI frequency; after removing number of methods from the models, SP and depressive symptoms emerged as most important in the prediction of NSSI frequency (Ammerman et al., 2017b).

5. Discussion

This systematic review provided a summary of studies utilizing ML techniques to advance the understanding and prediction of SITBs. The current review included findings from 35 articles, all published within the last 10 years. Based on this body of literature, we conclude that ML has demonstrated promise in significantly augmenting our prediction of SITBs. Despite observing a recent increase in the use of ML, we further conclude that these methods are still limited in their implementation in this field of research. Below, we aim to outline gaps in this literature and suggest ways to extend current findings in order to guide researchers to realize ML's potential to aid in the prediction and prevention of SITBs.

6. Review of research

In reviewing the findings of the included studies, it may be useful for us to consider each study as having one (or more) of three general aims in their use of ML analyses: (1) improving prediction accuracy, (2) identifying important model indicators (i.e., variable selection) and interactions, and (3) modeling underlying subgroups in the data. We briefly discuss studies in each area.

First, we turn to the identified studies that aimed to improve the prediction accuracy of SITBs. With previous meta-analyses finding our ability to predict suicidal and non-suicidal self-injurious behaviors to be near chance (e.g., weighted SA AUC = 0.58; weighted suicide death AUC = 0.57, weighted NSSI OR = 1.59; Franklin et al., 2017; Fox et al., 2015), we can see, even in the relatively small body of literature reviewed, that the use of ML techniques has offered improved prediction

over traditional statistical methodology in several studies (e.g., Walsh et al., 2017; Kessler et al., 2015; Kessler et al., 2017a, 2017b). Highlighting this, some authors have directly compared their ML findings to traditional methods, finding marked differences in SA prediction performance between methods (e.g., ML AUCs = 0.80–0.84 vs. multiple logistic regression AUCs = 0.66–0.68; Walsh et al., 2017) and instability in traditional methods (Kessler et al., 2017a, 2017b). The use of ML techniques for improved prediction may be most notable in predicting the outcome of suicide death as, despite its importance, limited research has focused on this low base rate outcome. Among the included reviewed studies, AUCs ranging from 0.71 to 0.89 were achieved in predicting suicide death (Kessler et al., 2015; Kessler et al., 2017a, 2017b). Importantly, improved model accuracy with the employment of ML for the prediction of suicidal behavior has been achieved in prediction windows as short as 7 days (Walsh et al., 2017) and as long as 2 years (Kessler et al., 2017a, 2017b), demonstrating its potential use in informing both crisis intervention and long-term prevention.

The aforementioned studies have demonstrated improved prediction predominantly through methods that permit minimal interpretability of individual variables (e.g., “black box” methods). While important to note the limitations with interpreting the influence of single indicators within ML models (e.g., Strobl et al., 2009), these methods, and others, have also been used for variable selection. Using ML for variable selection permits researchers to identify attributes from the data (indicators) that contribute to predictive model accuracy, ultimately allowing researchers to reduce the number of attributes in a model, simplifying models and increasing interpretability. Within the current review, studies that employed ML and commented on variable importance have served to replicate the findings of well-known predictors of future SITBs and SITB risk (e.g., depression, previous SITBs, psychiatric hospitalization; Bae et al., 2015; Hettige et al., 2017; Ilgen et al., 2009; Walsh et al., 2017), in addition to providing increased confidence in their importance in prediction as these factors are considered in conjunction with numerous other predictors and still emerge as important. Using ML for variable selection has also allowed researchers to identify several novel predictors from innovative data sources, many that have received relatively little attention in prior research. For example, it was found that the use of “frightened” and “agitated” within clinical notes were important for distinguishing those who died by suicide from psychiatric controls (Poulin et al., 2014). Similarly, in response to unstructured questions and social media post content the usage of specific words (Braithwaite et al., 2016; Cook et al., 2016) were found to differentiate between those with and without SITBs. Finally, even through the use of more traditional data sources, variables that have received relatively limited previous attention have been identified as important in SITB prediction: crime perpetration (Kessler et al., 2015; Kessler et al., 2017b), cocaine dependence (Passos et al., 2016), family intimacy (Bae et al., 2015), and specific NSSI characteristics (Burke et al., 2018). Through permitting the exploratory analysis of many variables simultaneously, ML can highlight novel variables in SITB prediction which may be important for inclusion in future research to improve model accuracy and, thus, SITB risk stratification.

Studies in the current review have also used ML to identify novel interactions, taking advantage of ML's ability to simultaneously consider a myriad of independent predictors and their interaction terms. Indeed, studies utilizing simple decision trees have identified particularly high-risk groups for specific SITB outcomes (e.g., Bae et al., 2015; Batterham & Christensen, 2012; Burke et al., 2018; Handley et al., 2014; Ilgen et al., 2009). Ilgen et al. (2009) found that veterans diagnosed with depression and a substance use disorder, who identify as non-African American, and who were recently psychiatrically hospitalized were at highest risk for death by suicide (Ilgen et al., 2009). Also finding a clinically relevant interaction, Bae et al. (2015) demonstrated that adolescents with high depression levels and who exhibited frequent delinquent behavior were at particularly high risk for past year SA and that this risk was even greater for females. Studies of this nature are particularly important given their high interpretability and, consequently, immediate clinical relevance in determining risk.

Finally, studies identified in the current review have utilized ML techniques to uncover potential subgroups in the data. While a limited number of studies used ML techniques for this purpose, the use of some ML methods (e.g., decision trees and extensions) have permitted the identification of optimal cutoffs for participant groupings based on a particular variable. For example, studies included in the current review aimed to examine how individuals who engage in NSSI may be alike, or may cluster, based on the relationship between NSSI characteristics (e.g., age of onset, number of acts) and other psychopathology variables (Ammerman et al., 2018; Ammerman et al., 2017a). Utilizing ML techniques in this fashion does not offer findings specific to prediction accuracy; however, results have the potential to empirically inform cutoffs or cut scores important in risk classification, the identification of unique risk factors, or diagnostic decision making and treatment planning.

The reviewed studies have demonstrated ML's promise in moving the field of SITB research forward through the concurrent examination of well-established and novel predictors. Further, the variety of ML implementation in the included studies highlights the numerous ways in which ML can be applied for not only improved prediction accuracy but also for variable (and interaction) selection and subgroup identification. Although the knowledge obtained thus far in the field has been immensely valuable, we now consider several possible directions to better leverage the advantages of ML to further advance SITB prediction and prevention.

7. Directions for future research

7.1. Broaden outcomes and indicators

The current systematic review highlights the need for researchers to extend both the outcomes and indicators included in our research. Importantly, only 35 papers using ML techniques to examine SITBs were identified; the majority of these studies were published within the past three years. This underscores the limited, but growing, body of research that has taken advantage of ML techniques and highlights the gaps in advancing SITB prediction and prevention. Specific outcomes have received particularly limited attention. Foremost, only five studies focused on suicide death as the outcome, all of which also utilized a military or veteran population. While an important, high-risk population of study (Kang et al., 2015; Kaplan et al., 2012), critical next steps in research include considering similar models in civilian populations to replicate results across samples, in addition to across settings (e.g., those already engaged in care, community samples), in order to inform broad implementations of suicide prevention interventions. Given that suicide death is a relatively rare occurrence, and consequentially involves resource-intensive data collection, work to expand findings from the current review may be best suited to leverage the use of big datasets, such as publicly available datasets of civilian samples (e.g., National Death Index), which given their size, are particularly well-suited for the implementation of ML techniques. Similarly, a small number of studies have focused solely on SP (one study) and NSSI (three studies). Beyond the distress and impairment associated with the occurrence of these experiences (e.g., Mars et al., 2014), approximately 55–70% of those having a SP go on to attempt suicide (Kessler et al., 1999; Nock et al., 2008) and those with a NSSI history are at four times the risk of SA (Ribeiro et al., 2016), making these two outcomes important areas of study with broad implications for suicide prevention efforts. Furthermore, an important next step for future research is including multiple outcomes within one study for improved identification of factors that facilitate the progression from thoughts to behavior. Only a few studies identified in the current review did this (e.g., Burke et al.,

2018; Just et al., 2017); results suggested importance of indicators was in fact dependent on outcome.

Although several studies (e.g., Oh et al., 2017; Kessler et al., 2017b, 2017b; Walsh et al., 2017) in the current review employed varying prediction time horizons, no identified studies can inform the prediction of *imminent* risk for SITBs. Emphasizing the importance of imminent risk prediction is work highlighting the instability of the occurrence of SITB outcomes (e.g., Kleiman et al., 2017), and, as demonstrated in the current review, the changing importance of risk factors depending on prediction window (720 versus 7 days prior to suicide attempt; Walsh et al., 2017). Moving forward, it will be necessary for researchers to consider constructs at more precise levels of measurement (i.e., daily, hourly), in addition to the interactions of dynamic factors with more static predictors (i.e., previous suicide attempts, history of psychiatric hospitalization). As time-intensive data (e.g., ecological momentary assessment) becomes more accessible and more technologically advanced and robust (e.g., physiological wearables), ML has promise in leveraging the resultant large datasets to identify risk factors in the hours and days prior to SITBs.

In addition to increasing attention toward understudied outcomes, and at differing prediction windows, it will be important for future research to continue increasing the array of indicators used in models. For example, given the high comorbidity of affective disorders and suicide outcomes (Nock et al., 2009), and the importance of affective disorders in predicting suicidal behaviors (Franklin et al., 2017), it may be important for future research to consider this association using ML techniques, which may be particularly appropriate given the high correlations between variables and the likely importance of variable interactions. Several of the studies included in the current review also utilized non-traditional data (e.g., Braithwaite et al., 2016; Pestian et al., 2017) in ML models, however there is still room for expansion. With the growing integration of electronic communications, both professionally (i.e., secure messaging with healthcare providers) and personally (i.e., social media applications), ML techniques could be used to harness the massive amount of existing text and image data, reducing reliance on individual self-reports, to help identify potentially suicidal or distressed individuals. For example, surveillance of pervasive social media platforms (e.g., Facebook, Twitter) using large-scale ML-based algorithms, could improve identification of at-risk populations that may have increased barriers to care (e.g., due to geographical location, stigma, physical ailments). Other forms of data that are relatively untapped, but may ultimately aid in suicide risk identification, include passive phone data (e.g., GPS, texting, call logs, web search histories), existing public databases (legal, financial, etc.; Kessler et al., 2017a, 2017b), and qualitative responses to unstructured questions (e.g., Pestian et al., 2016). The potential of these data structures is highlighted in the current review through the utilization of social media (e.g., Braithwaite et al., 2016; Cheng et al., 2017; Guan et al., 2015), clinical note text (Poulin et al., 2014), and verbal and nonverbal responses to unstructured questions (Pestian et al., 2016; Pestian et al, 2017), each producing innovative findings.

7.2. Broadening research questions

One of the advantages of ML techniques is the ability to consider novel research questions, many of which cannot be easily addressed by traditional statistical methods. One branch of ML techniques that may be particularly suited for this extension, and has yet to be employed in SITB research, is unsupervised learning. Unsupervised learning may be used to identify natural groupings or clusters of individuals. For example, studies may identify data-driven phenotypes of suicide risk based on numerous self-report measures, diagnostic procedures, cognitive indices, and biomarkers. Unsupervised learning also holds promise in being able to cluster SITB trajectories, which may allow more nuanced investigations of longitudinal effects such as the temporal transition from suicidal thoughts to suicidal behavior.

Research questions related to diagnostic considerations, risk classification, and treatment planning may also be supported through the implementation of ML. Compared to traditional analytic methods, specific ML techniques (e.g., decision trees, structural equation modeling trees) are well-suited to simultaneously examining multiple risk factors (in addition to the interactions between variables) in order to provide meaningful splits in the data that are needed to improve our understanding of potential subgroups of individuals with SITB (e.g., treatment responding subgroups, diagnostic subgroups).

7.3. Advancing ML techniques

A variety of ML methods were applied across a number of studies. Given the varying types of datasets used, along with different metrics reported, it is difficult to reach an over-arching conclusion about model performance across studies. Although accuracy was not generally high (e.g., most reported < 0.90), comparing reported accuracy across studies is difficult because this metric depends on the distribution of the outcome. For example, this metric requires taking into account how well we could accurately predict both outcomes (e.g., the presence and absence of a SA) beyond classifying all cases as negative (e.g., not having a SA). Particularly for imbalanced outcomes (e.g., a majority of respondents had no history of SA), there exist alternative metrics that give us a more realistic picture of a model's performance (see Saito and Rehmsmeier, 2015 for discussion on precision-recall curves), and sampling schemes to improve prediction (see Chawla, 2009). We also look forward to the application of ML methods to novel forms of data collection, allowing for the prediction of SITB across a wider variety of time horizons, with hopes this spurns the development of newer ML methods that specifically incorporate the longitudinal nature of the data, as well the unique challenges of predicting SITB. Finally, in addition to increasing the number of ML techniques for longitudinal assessment, we encourage broadening the use and development of ML methods that incorporate measurement error. Particularly when the data quality is not high (e.g., large amounts of measurement noise in the predictors or outcomes, outliers, high amounts of missing data), or the sample size is not large, simpler methods like regularized regression will often perform comparably to methods that incorporate nonlinearity. Data quality often poses constraints on the amount of information that can be extracted, thus limiting the complexity of ML methods that are needed to model the relationships. Indeed, our limitations in predicting suicide are not only a feature of the outcome, but also the measurement error of most predictors in psychological datasets. Extending ML techniques, such as regularization and decision trees to the structural equation modeling framework (see Jacobucci et al., 2018 or Brandmaier et al., 2013), allows for a combination of benefits of using latent variables and aspects of ML, hopefully increasing our prediction accuracy.

8. Limitations

The limitations of this systematic review should be addressed. First, we identified only 35 empirical studies that met our inclusion criteria and, as a result, the conclusions of this review should be interpreted cautiously given the small sample size, particularly when considering sample size by SITB outcome. Second, the majority of these studies employed cross-sectional designs (n = 25/35; 71.42%), limiting the extent to which prospective conclusions may be drawn. Third, the use of inconsistent reporting methods on classifier performance made it impractical to compare model performance across studies. As such, we did not provide commentary on relative performance of models (or a corresponding meta-analytic examination of findings). As a greater number of studies are conducted, future review papers would do well to compare model performance, taking into account the marked differences in design, such as number of indicators and type(s) of ML technique used. Finally, a common suicide prediction-specific limitation

that applies to both traditional and ML statistical approaches is the problem of class imbalance. Although reporting the AUC provides a better understanding of model performance than accuracy in the case of class imbalance, this metric can still misrepresent results. Despite it helping to give a sense of how well an algorithm performs in classifying both positive and negative cases, a high AUC in the context of a high class imbalance can merely reflect a strong ability to predict negative cases. In order to better reflect model performance in the case of class imbalance, some of the studies in our review reported precision-recall statistics, which better reflect the performance of a model in predicting positive cases. Whereas some studies addressed class imbalance in their *reporting* of model performance, many fewer studies discuss whether class imbalance was addressed, beyond the use of CV, in their sampling strategy (e.g., Kessler et al., 2017b; Mann et al., 2008; Passos et al., 2016) or in training the classifier (e.g., Kuroki and Tilley, 2012; Passos et al. 2016). Both sampling strategies, which select subsets of majority class or bootstrap the minority class to create more equal class distributions when training the classifier, and class weighting (i.e. incurring a larger cost for misclassifying minority class members) are strategies for improving the model prediction performance, which go above and beyond just the use of different performance metrics for more informative model performance assessment.

The more general limitations of machine learning methods should also be considered. Although ML approaches are often more resistant to over-fitting as compared to traditional approaches, ML approaches remain vulnerable to over-fitting data (i.e., creating a model that fits the training data well yet has minimal ability to classify cases accurately in a separate sample). Despite the fact that many of the studies in this review implemented CV methods (e.g., testing algorithms on holdout samples), few studies validated their algorithms on external samples. Without access to external samples to conduct a truly independent test of a model's performance, we are unable to ascertain the full extent to which over-fitting may be occurring. A related and important limitation to consider when interpreting the ML models presented in this review is that the predictors identified as 'important' in the ML models are based on increasing within-sample prediction accuracy; such 'important' factors may emerge due to over-fitting and thus may not be generalizable to other samples (Strobl et al., 2009). These limitations further corroborate the need for caution in interpreting findings both across and within studies. There is a need for proof of concept studies to move the field forward, where future studies with larger sample sizes and access to corresponding external data sets may be able to replicate and extend the current findings.

9. Implications and conclusions

The current review highlights the potential for significant clinical advancements through the use of ML in predicting SITBs. Several studies identified within the review (e.g., Kessler et al., 2015, 2017a, 2017b; Walsh et al., 2017) demonstrated markedly improved prediction of suicidal behavior by utilizing data existing in patient EMRs. However, given that this body of literature has only recently emerged and is still quite limited (i.e., 35 studies, general lack of external replication), we would hesitate to suggest alterations in the design of public policies at this stage. However, if more studies are conducted that exhibit increasingly strong prediction, and if these studies are able to prove model performance in external samples, then public policies may consider implementing machine learning approaches to enhance in-vivo monitoring of suicide risk at a public health level.

The findings presented may have immediate implications for the large-scale implementation of ML techniques in healthcare settings that aim to test the feasibility of these approaches. Indeed, healthcare systems (e.g., Office of Public and Intergovernmental Affairs in Veterans Health Administration, 2017) have already begun to use ML methods to improve risk stratification by harnessing the value of ML as applied to EMRs. This approach allows for the leveraging of ML in large datasets while balancing the need for clinically relevant outcomes (i.e., a targeted group for prevention efforts); a comparable methodology may prove useful in similar healthcare systems or potentially even in mental health clinics. We also speculate there is promise for further improved predictive power when utilizing richer indicators, such as objective features (e.g., implicit attitudes about suicide, linguistic features of patient verbalizations) in conjunction with EMR data. This represents a fruitful area for future research and clinical endeavors. Beyond the identification of those at risk, the use of ML can provide valuable information directly related to clinical treatment planning. For example, these techniques can aid in identifying empirically supported cutoff scores directly relevant to the development of empirically-supported diagnostic categories (e.g., Ammerman et al., 2017a). The use of ML may be effective in revising and modifying suicide risk assessment in a way that allows for the integration of numerous risk factors to determine which may be most important to capture high risk, an approach that will aid in quick and accurate risk determinations as the emphasis on suicide risk screening grows (e.g., The Joint Commission, 2016).

In addition to implications for individual clinical care, ML may also be pertinent at a population level. For example, Facebook and Instagram have begun using ML pattern recognition techniques to detect posts with suicide-related material and provide resources to the person who posted the content (e.g., crisis line number, ability to message with a crisis worker; CNN, 2016; Facebook, 2017). The implementation of ML techniques into large-scale, continually updating internet-based databases represent low-cost models that may improve suicide risk assessment and inform suicide prevention efforts. Indeed, to take this one step further, unsupervised ML may also be able to guide treatment decision making by finding subgroups that may respond to inexpensive evidence-based therapies (e.g., internet-delivered), which can augment prevention for the numerous individuals that may be identified through large-scale prediction, and in turn save limited resources for those necessitating more comprehensive or experimental approaches (Kessler et al., 2017a, 2017b). These ideas underscore the broader impact that ML may have on suicide prevention.

Overall, ML techniques have already made a significant advancement in suicide prediction, despite their limited application. This area of research represents one ripe for significant growth not only in our prediction of suicide related events, but also in the appropriate implementation of prevention and intervention efforts, which has positive implications for the reduction in suicide rates overall.

Conflicts of interest

None.

Acknowledgements

This research was supported by a University of Michigan James N. Morgan Fund grant to Taylor A. Burke.

The funding source had no role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Contributors: Taylor A. Burke, Brooke Ammerman, and Ross Jacobucci contributed to the study design, data collection, analyses, and manuscript preparation.

References

Ammerman, B.A., Jacobucci, R., Kleiman, E.M., Muehlenkamp, J.J., McCloskey, M.S., 2017a. Development and validation of empirically derived frequency criteria for NSSI disorder using exploratory data mining. Psychol. Assess. 29, 221–231. https://doi.org/10.1037/pas0000334.

Ammerman, B.A., Jacobucci, R., Kleiman, E.M., Uyeji, L.L., McCloskey, M.S., 2018. The relationship between nonsuicidal self-injury age of onset and severity of self-harm. Suicide Life-Threatening Behav. 48, 31–37. https://doi.org/10.1111/sltb.12330.

Ammerman, B.A., Jacobucci, R., McCloskey, M.S., 2017b. Using exploratory data mining

to identify important correlates of nonsuicidal self-injury frequency. Psychol. Violence. https://doi.org/10.1037/vio0000146.
Baca-Garcia, E., Vaquero-Lorenzo, C., Perez-Rodriguez, M.M., Gratacòs, M., Bayés, M., Santiago-Mozos, R., Leiva-Murillo, J.M., De Prado-Cumplido, M., Artes-Rodriguez, A., Ceverino, A., Diaz-Sastre, C., Fernandez-Navarro, P., Costas, J., Fernandez-Piqueras, J., Diaz-Hernandez, M., De Leon, J., Baca-Baldomero, E., Saiz-Ruiz, J., Mann, J.J., Parsey, R.V., Carracedo, A., Estivill, X., Oquendo, M.A., 2010. Nucleotide variation in central nervous system genes among male suicide attempters. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. https://doi.org/10.1002/ajmg.b.30975.
Bae, S.M., Lee, S.A., Lee, S.-H., 2015. Prediction by data mining, of suicide attempts in Korean adolescents: a national study. Neuropsychiatr. Dis. Treat. 11 (2015) SepArtID 2367-2375 11.
Barros, J., Morales, S., Echávarri, O., García, A., Ortega, J., Asahi, T., Moya, C., Fischman, R., Maino, M.P., Núñez, C., 2017. Suicide detection in Chile: Proposing a predictive model for suicide risk in a clinical sample of patients with mood disorders. Rev. Bras. Psiquiatr. 39, 1–11. https://doi.org/10.1590/1516-4446-2015-1877.
Batterham, P.J., Christensen, H., 2012. Longitudinal risk profiling for suicidal thoughts and behaviours in a community cohort using decision trees. J. Affect. Disord. https://doi.org/10.1016/j.jad.2012.05.021.
Braithwaite, S.R., Giraud-Carrier, C., West, J., Barnes, M.D., Hanson, C.L., 2016. Validating machine learning algorithms for twitter data against established measures of suicidality. JMIR Ment. Heal. 3, e21. https://doi.org/10.2196/mental.4822.
Brandmaier, A.M., von Oertzen, T., Mcardle, J.J., Lindenberger, U., 2013. Structural equation model trees. Psychol. Methods. https://doi.org/10.1037/a0030001.
Burke, T.A., Jacobucci, R., Ammerman, B.A., Piccirillo, M., McCloskey, M.S., Heimberg, R.G., Alloy, L.B., 2018. Identifying the relative importance of non-suicidal self-injury features in classifying suicidal ideation, plans, and behavior using exploratory data mining. Psychiatry Res. 262. https://doi.org/10.1016/j.psychres.2018.01.045.
Center for Disease Control and Prevention (CDC), 2016. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control [WWW Document]. Webbased Inj. Stat. Query Report. Syst.
Chawla, N.V., 2009. Data mining for imbalanced datasets: an overview. Data Mining and Knowledge Discovery Handbook. . https://doi.org/10.1007/978-0-387-09823-4_45.
Cheng, Q., Li, T.M., Kwok, C.L., Zhu, T., Yip, P.S., 2017. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. J. Med. Internet Res. 19, 1–15. https://doi.org/10.2196/jmir.7276.
CNN, 2016. Instagram launches suicide prevention tool. http://money.cnn.com/2016/10/20/technology/instagram-suicide-prevention-tools/index.html, Accessed date: 23 May 2018.
Cook, B.L., Progovac, A.M., Chen, P., Mullin, B., Hou, S., Baca-Garcia, E., 2016. Novel Use of Natural Language Processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput. Math. Methods Med. 2016. https://doi.org/10.1155/2016/8708434.
Crosby, A., Ortega, L., Melanson, C., 2011. Self-directed Violence surveillance: Uniform definitions and Recommended Data elements, Version 1.0 (RPRT). Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Atlanta, GA.
Curtin, S.C., Warner, M., Hedegaard, H., 2016. Increase in Suicide in the United States, 1999-2014. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
Delgado-Gomez, D., Baca-Garcia, E., Aguado, D., Courtet, P., Lopez-Castroman, J., 2016. Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. J. Affect. Disord. 206, 204–209. https://doi.org/10.1016/j.jad.2016.07.032.
Facebook, 2017. Building a safer community with new suicide prevention tools. https://newsroom.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/, Accessed date: 23 May 2018.
Fox, K.R., Franklin, J.C., Ribeiro, J.D., Kleiman, E.M., Bentley, K.H., Nock, M.K., 2015. Meta-analysis of risk factors for nonsuicidal self-injury. Clin. Psychol. Rev. https://doi.org/10.1016/j.cpr.2015.09.002.
Franklin, J.C., Ribeiro, J.D., Fox, K.R., Bentley, K.H., Kleiman, E.M., Huang, X., Musacchio, K.M., Jaroszewski, A.C., Chang, B.P., Nock, M.K., 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol. Bull. https://doi.org/10.1037/bul0000084.
Gradus, J.L., King, M.W., Galatzer-Levy, I., Street, A.E., 2017. Gender differences in machine learning models of trauma and suicidal ideation in veterans of the Iraq and Afghanistan wars. J. Trauma. Stress. https://doi.org/10.1002/jts.22210.
Guan, L., Hao, B., Cheng, Q., Yip, P.S., Zhu, T., 2015. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. JMIR Ment. Heal. 2, e17. https://doi.org/10.2196/mental.4227.
Handley, T.E., Hiles, S.A., Inder, K.J., Kay-Lambkin, F.J., Kelly, B.J., Lewin, T.J., McEvoy, M., Peel, R., Attia, J.R., 2014. Predictors of suicidal ideation in older people: a decision tree analysis. Am. J. Geriatr. Psychiatry. https://doi.org/10.1016/j.jagp.2013.05.009.
Hettige, N.C., Nguyen, T.B., Yuan, C., Rajakulendran, T., Baddour, J., Bhagwat, N., Bani-Fatemi, A., Voineskos, A.N., Mallar Chakravarty, M., De Luca, V., 2017. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: a machine learning approach. Gen. Hosp. Psychiatry 47, 20–28. https://doi.org/10.1016/j.genhosppsych.2017.03.001.
Hill, R.M., Oosterhoff, B., Kaplow, J.B., 2017. Prospective identification of adolescent suicide ideation using classification tree analysis: models for community-based screening. J. Consult. Clin. Psychol. 85, 702–711. https://doi.org/10.1037/ccp0000218.
Ilgen, M.A., Downing, K., Zivin, K., Hoggatt, K.J., Kim, H.M., Ganoczy, D., Austin, K.L., McCarthy, J.F., Patel, J.M., Valenstein, M., 2009. Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. J. Clin. Psychiatry 70, 1495–1500. https://doi.org/10.4088/JCP.08m04795.
Jacobucci, R., Kievit, R., Brandmaier, A.M. (2018, February 16). Variable selection in structural equation models with regularized MIMIC models. Retrieved from osf.io/z2dtq.
Jordan, P., Shedden-Mora, M.C., Löwe, B., 2018. Predicting suicidal ideation in primary care: an approach to identify easily assessable key variables. Gen. Hosp. Psychiatry 51, 106–111. https://doi.org/10.1016/j.genhosppsych.2018.02.002.
Just, M.A., Pan, L., Cherkassky, V.L., McMakin, D.L., Cha, C., Nock, M.K., Brent, D., 2017. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. Nat. Hum. Behav. 1, 911–919. https://doi.org/10.1038/s41562-017-0234-y.
Kang, H.K., Bullman, T.A., Smolenski, D.J., Skopp, N.A., Gahm, G.A., Reger, M.A., 2015. Suicide risk among 1.3 million veterans who were on active duty during the Iraq and Afghanistan wars. Ann. Epidemiol. https://doi.org/10.1016/j.annepidem.2014.11.020.
Kaplan, M.S., McFarland, B.H., Huguet, N., Newsom, J.T., 2012. Estimating the risk of suicide among US veterans: how should we proceed from here? Am. J. Public Health. https://doi.org/10.2105/AJPH.2011.300611.
Kessler, R.C., Borges, G., Walters, E.E., 1999. Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.56.7.617.
Kessler, R.C., Hwang, I., Hoffmire, C.A., McCarthy, J.F., Petukhova, M.V., Rosellini, A.J., Sampson, N.A., Schneider, A.L., Bradley, P.A., Katz, I.R., Thompson, C., Bossarte, R.M., 2017a. Developing a practical suicide risk prediction model for targeting high-risk patients in the veterans health administration. Int. J. Methods Psychiatr. Res. 26, 1–7. https://doi.org/10.1002/mpr.1575.
Kessler, R.C., Stein, M.B., Petukhova, M.V., Bliese, P., Bossarte, R.M., Bromet, E.J., Fullerton, C.S., Gilman, S.E., Ivany, C., Lewandowski-Romps, L., Millikan Bell, A., Naifeh, J.A., Nock, M.K., Reis, B.Y., Rosellini, A.J., Sampson, N.A., Zaslavsky, A.M., Ursano, R.J., 2017b. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Mol. Psychiatry. https://doi.org/10.1038/mp.2016.110.
Kessler, R.C., Warner, C.H., Ivany, C., Petukhova, M.V., Rose, S., Bromet, E.J., Brown, M., Cai, T., Colpe, L.J., Cox, K.L., Fullerton, C.S., Gilman, S.E., Gruber, M.J., Heeringa, S.G., Lewandowski-Romps, L., Li, J., Millikan-Bell, A.M., Naifeh, J.A., Nock, M.K., Rosellini, A.J., Sampson, N.A., Schoenbaum, M., Stein, M.B., Wessely, S., Zaslavsky, A.M., Ursano, R.J., 2015. Predicting suicides after psychiatric hospitalization in US army soldiers: The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). JAMA Psychiatry 72, 49–57. https://doi.org/10.1001/jamapsychiatry.2014.1754.
Kleiman, E.M., Turner, B.J., Fedor, S., Beale, E.E., Huffman, J.C., Nock, M.K., 2017. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J. Abnorm. Psychol. https://doi.org/10.1037/abn0000273.
Kuroki, Y., 2015. Risk factors for suicidal behaviors among Filipino Americans: a data mining approach. Am. J. Orthopsychiatry 85, 34–42. https://doi.org/10.1037/ort0000018.
Kuroki, Y., Tilley, J.L., 2012. Recursive partitioning analysis of lifetime suicidal behaviors in Asian Americans. Asian Am. J. Psychol. https://doi.org/10.1037/a0026586.
Lopez-Castroman, J., Perez-Rodriguez, M., de las, M., Jaussent, I., Alegria, A.A., Artes-Rodriguez, A., Freed, P., Guillaume, S., Jollant, F., Leiva-Murillo, J.M., Malafosse, A., Oquendo, M.A., de Prado-Cumplido, M., Saiz-Ruiz, J., Baca-Garcia, E., Courtet, P., Perroud, N., Saiz, P.A., Baca-Garcia, E., LopezCastroman, J., Blasco-Fontecilla, H., Sarchiapone, M., Carli, V., Courtet, P., Jaussent, I., Guillaume, S., Malafosse, A., 2011. Distinguishing the relevant features of frequent suicide attempters. J. Psychiatr. Res. 45, 619–625. https://doi.org/10.1016/j.jpsychires.2010.09.017.
Mann, J.J., Ellis, S.P., Waternaux, C.M., Liu, X., Oquendo, M.A., Malone, K.M., Brodsky, B.S., Haas, G.L., Currier, D., 2008. Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making. J. Clin. Psychiatry. https://doi.org/10.4088/JCP.v69n0104.
Mars, B., Heron, J., Crane, C., Hawton, K., Lewis, G., Macleod, J., …, Gunnell, D., 2014. Clinical and social outcomes of adolescent self harm: population based birth cohort study. BMJ 349, g5954.
[Ed]McArdle, J.J., Ritschard, G., 2014. Contemporary Issues In Exploratory Data Mining in the Behavioral Sciences.
Metzger, M.-H., Tvardik, N., Gicquel, Q., Bouvry, C., Poulet, E., Potinet-Pagliaroli, V., 2017. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. Int. J. Methods Psychiatr. Res. 26, e1522. https://doi.org/10.1002/mpr.1522.
Morales, S., Barros, J., Echávarri, O., García, F., Osses, A., Moya, C., Maino, M.P., Fischman, R., Núñez, C., Szmulewicz, T., Tomicic, A., 2017. Acute mental discomfort associated with suicide behavior in a clinical sample of patients with affective disorders: ascertaining critical variables using artificial intelligence tools. Front. Psychiatry 8, 1–16. https://doi.org/10.3389/fpsyt.2017.00007.
Morgan, J.N., 2005. History and potential of binary segmentation for exploratory data analysis. J. Data Sci. 3, 123–136.
Nock, M.K., 2009. Why do people hurt themselves?: New insights into the nature and functions of self-injury. Curr. Dir. Psychol. Sci. https://doi.org/10.1111/j.1467-8721.2009.01613.x.
Nock, M.K., Borges, G., Bromet, E.J., Alonso, J., Angermeyer, M., Beautrais, A., Bruffaerts, R., Chiu, W.T., de Girolamo, G., Gluzman, S., de Graaf, R., Gureje, O., Haro, J.M., Huang, Y., Karam, E., Kessler, R.C., Lepine, J.P., Levinson, D., Medina-Mora, M.E., Ono, Y., Posada-Villa, J., Williams, D., 2008. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br. J. Psychiatry. https://doi.org/10.1192/bjp.bp.107.040113.
Nock, M.K., Hwang, I., Sampson, N., Kessler, R.C., Angermeyer, M., Beautrais, A., Borges,

G., Bromet, E., Bruffaerts, R., De Girolamo, G., De Graff, R., 2009. Cross-national analysis of the associations among mental health disorders and suicidal behavior: findings from the WHO World Mental Health Surveys. PLOS Med. https://doi.org/10.1371/journal.pmed.1000123.

Office of Public and Intergovernmental Affairs in Veterans Health Administration, 2017. VA REACH VET initiative helps save veterans lives: program signals when more help is needed for at-risk veterans. https://www.va.gov/opa/pressrel/pressrelease.cfm?id=2878 12 May 2017.

Oh, J., Yun, K., Hwang, J.-H., Chae, J.-H., 2017. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front. Psychiatry 8, 1–8. https://doi.org/10.3389/fpsyt.2017.00192.

Pestian, J.P., Grupp-Phelan, J., Bretonnel Cohen, K., Meyers, G., Richey, L.A., Matykiewicz, P., Sorter, M.T., 2016. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide Life-Threatening Behav. 46, 154–159. https://doi.org/10.1111/sltb.12180.

Passos, I.C., Mwangi, B., Cao, B., Hamilton, J.E., Wu, M.J., Zhang, X.Y., Zunta-Soares, G.B., Quevedo, J., Kauer-Sant'Anna, M., Kapczinski, F., Soares, J.C., 2016. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. J. Affect. Disord 193, 109–116. https://doi.org/10.1016/j.jad.2015.12.06.

Pestian, J.P., Sorter, M., Connolly, B., Bretonnel Cohen, K., McCullumsmith, C., Gee, J.T., Morency, L.P., Scherer, S., Rohlfs, L., Faist, R., Korbee, L., McCord, L., McCourt, R., Murphy, C., 2017. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide Life-Threatening Behav 47, 112–121. https://doi.org/10.1111/sltb.12312.

Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., Watts, B., Flashman, L., McAllister, T., 2014. Predicting the risk of suicide by analyzing the text of clinical notes. PLoS One 9, 1–8. https://doi.org/10.1371/journal.pone.0085733.

Ribeiro, J.D., Franklin, J.C., Fox, K.R., Bentley, K.H., Kleiman, E.M., Chang, B.P., Nock, M.K., 2016. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol. Med. https://doi.org/10.1017/S0033291715001804.

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. https://doi.org/10.1371/journal.pone.0118432.

Shepard, D.S., Gurewich, D., Lwin, A.K., Reed, G.A., Silverman, M.M., 2016. Suicide and suicidal attempts in the United States: costs and policy implications. Suicide Life. Threat. Behav. https://doi.org/10.1111/sltb.12225.

Steyerberg, E.W., Harrell, F.E., 2016. Prediction models need appropriate internal, internal-external, and external validation. J. Clin. Epidemiol. https://doi.org/10.1016/j.jclinepi.2015.04.005.

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods. https://doi.org/10.1037/a0016973.

The Joint Commission, 2016. Detecting and treating suicide ideation in all settings. Sentinel Event Alert Retrieved from. https://www.jointcommission.org/assets/1/18/SEA_56_Suicide.pdf.

Walsh, C.G., Ribeiro, J.D., Franklin, J.C., 2017. Predicting risk of suicide attempts over time through machine learning. Clin. Psychol. Sci. 5, 457–469. https://doi.org/10.1177/2167702617691560.

World Health Organization, 2014. Preventing Suicide: A Global Imperative. World Health Organization. http://www.who.int/mental_health/suicide-prevention/world_report_2014/en/.