Data mining of hospital suicidal and self-harm presentation records using a tailored evolutionary algorithm

Stapelberg, Nicolas; Randall, Marcus; Sveticic, Jerneja; Fugelli, Peter; Dave, Hasmeera; Turner, Kathryn

Published in:
Machine Learning with Applications

DOI:
10.1016/j.mlwa.2020.100012

E-pub ahead of print: 15/03/2021

Machine Learning with Applications 3 (2021) 100012

Contents lists available at ScienceDirect

Machine Learning with Applications

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

Data mining of hospital suicidal and self-harm presentation records using a tailored evolutionary algorithm

Nicolas J.C. Stapelberg [a,b,*], Marcus Randall [c], Jerneja Sveticic [a], Pete Fugelli [d], Hasmeera Dave [b], Kathryn Turner [a]

[a] Gold Coast Mental Health and Specialist Services, Gold Coast Health, Queensland, Australia
[b] Faculty of Health Science and Medicine, Bond University Queensland, Australia
[c] Bond Business School, Bond University, Queensland, Australia
[d] Healthcare Improvement Unit, Clinical Excellence Division, Department of Health, Queensland, Australia

A R T I C L E I N F O

Keywords:
Suicide prevention
Evolutionary algorithms
Machine learning
Data mining
SERoSP
n-grams

A B S T R A C T

The assessment of outcomes for the Gold Coast Mental Health and Specialist Services Suicide Prevention Strategy implementation required data on suicidal and self-harm presentations to be captured from the Emergency Department Information System (EDIS) database. Suicidal and self-harm presentations are not uniformly coded in the EDIS and require human assessment to differentiate these presentations from other cases (e.g., accidental injuries). A novel evolutionary algorithm was used to learn weighting variables from a psychiatrist-rated training dataset in order to generate an appropriate cut-off score for identifying suicidal and self-harm presentations from EDIS. The resulting *Searching EDIS for Records of Suicidal Presentations (SERoSP)* program was then run on a psychiatrist-rated validation dataset using the weights generated by the algorithm. SERoSP is optimised to be able to detect suicidal and self-harm presentations with a high degree of accuracy (a sensitivity of 0.95 and a specificity of 0.92). The SERoSP program is a reliable and cost-effective tool for the identification of suicidal and self-harm presentations from EDIS data, and is currently being successfully used in the suicide prevention strategy evaluation.

1. Introduction

Suicide is a complex, global, biopsychosocial phenomenon with a significant human toll. Based on World Health Organization estimates there were approximately 800,000 suicides globally in 2015, which equals a suicide rate of 10.7 per 100,000 annually (World Health Organization, 2017). In 2017, 3128 people died by suicide in Australia, almost 9 people per day (Australian Bureau of Statistics, 2016). The Zero Suicide Framework (ZSF) followed from the Suicide Care in Systems Framework (Covington et al., 2011) and 2012 US National Strategy for Suicide Prevention (Office of the Surgeon General (US and National Action Alliance for Suicide Prevention), 2012). The Henry Ford Health System initiated the Perfect Depression Care project and put forward that the aspirational goal of zero suicides as an essential component that helped to drive a dramatic and sustained reduction in suicides in their system (Coffey & Coffey, 2016) and the Zero suicide concept has been increasingly widely embraced, e.g., the Mersey Care Zero Suicide Policy (Mersey Care, 2015). The ZSF is now being adopted by multiple hospital and health services in Queensland, Australia, as part of a multi-site collaborative effort.

The Gold Coast Mental Health and Specialist Services Suicide Prevention Strategy (GCMHSS SPS) is the largest clinical implementation of a ZSF in Australia. Roll-out of the Suicide Prevention Pathway, a clinical component of the GCMHSS SPS, commenced in December 2016, along with a framework for process and outcomes evaluation.

Being able to accurately identify suicidal presentations from hospital records is of vital importance. With large volumes of data, automated methods of data mining become important tools. However, there is substantial heterogeneity in how the data collected by hospital and health services is coded. This paper describes the development of a new classification oriented Evolutionary Algorithm (EA) that produces weighting factors to predict whether a consumer record represents a suicidal presentation or not. Searching EDIS for Records of Suicidal Presentations (SERoSP) is a tool for data mining which has been trained using an evolutionary algorithm. Its purpose is to identify suicidal presentations in the Emergency Department Information System (EDIS)

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author at: Gold Coast Mental Health and Specialist Services, Gold Coast Health, Queensland, Australia.
E-mail addresses: Chris.Stapelberg@health.qld.gov.au (N.J.C. Stapelberg), mrandall@bond.edu.au (M. Randall), jerneja.sveticic@health.qld.gov.au (J. Sveticic), pete.fugelli@health.qld.gov.au (P. Fugelli), hasmeera.dave@student.bond.edu.au (H. Dave), kathryn.turner@health.qld.gov.au (K. Turner).

https://doi.org/10.1016/j.mlwa.2020.100012
Received 7 June 2020; Received in revised form 17 November 2020; Accepted 18 November 2020
Available online 25 November 2020

database, used in hospital and health services in Queensland, Australia, with a high degree of accuracy. The challenge in this task lies in the heterogeneity of coding into the EDIS database, as we describe in Section 3. The rest of this paper is organised as follows. Section 2 provides background on suicide prevention and existing EAs that have been applied to this area. Section 3 describes the problem addressed in this work and presents the Searching EDIS for Records of Suicidal Presentations (SERoSP) framework. Section 4 describes the methodology and results of the EA. Finally, Sections 5 and 6 comprise a discussion of the results and the broader implications of the work, as well as recommendations for future work that will arise from of this project.

2. The use of computational based methods in suicide-related research

Data mining – a large component of machine learning – can be defined as a search for structure and patterns in data using algorithmic methods (Baca-García et al., 2006; Goodwin et al., 2003). This field of data science deals with information technology systems that ‘learn’ from experience, observation, or other means, and is increasingly applied to high-volume, high throughput medical data (Oquendo et al., 2012). Joubert et al. (2012) defined clinical data mining as “a practise-based research strategy for systematically collecting and retrospectively analysing existing agency data to answer research questions and evaluate practise interventions” (p. 68).

EAs, which include genetic algorithms, constitute an important branch of machine learning. A complex problem might have many solutions and EAs provide powerful means to solve optimisation, modelling, simulation and search problems in multiple disciplines (Eiben et al., 2003). EAs have been employed broadly in the field of mental health, for example to examine the spatial distribution of mental disorders in Barcelona (Salinas-Pérez et al., 2015), or to compare characteristics of mental health patients visiting the psychiatric emergency department according to their time of arrival (Carmona et al., 2011).

Data mining is also increasingly used to study complex data on suicidal and self-harm presentations, particularly as suicide has been recognised as a public health problem that “challenges prediction due to its transdiagnostic, yet rare occurrence at the population-level” (Bernert et al., 2020) (p.1). A systematic review by Burke et al. (2019) identified 35 articles on the application of machine learning techniques to predict suicidal thoughts and behaviours published until February 2018. They reported three broad categories of the applicability of machine learning analyses for suicide prevention: (1) to improve prediction accuracy, (2) to identify important risk factors and interactions between them, and (3) to better describe underlying subgroups. More recently, a systematic review of the use of artificial intelligence and suicide prevention (Bernert et al., 2020) identified 87 reports that examined risk for suicidal ideation, suicide attempts, or death by suicide through the use of machine learning methods. Despite considerable diagnostic and methodological heterogeneity, high levels of model performance were observed, with machine learning-guided risk stratification models outperforming those relying on clinician-based prediction. Below, we summarise main findings from some of the most prominent publications from the growing literature in this field.

A review of machine learning methods and applications for detection of suicidal ideation notes that specific applications of this work cover a range of domains including questionnaires, suicide notes, and online user content, and electronic health records (Ji et al., 2019). The latter were used in an Australian study (Nguyen et al., 2016) which demonstrated the robustness and superior efficacy of three randomised machine learning techniques which efficiently manage high dimensionality and redundancy (random forests, gradient boosting machines, and deep neural nets with dropout) in predicting short and medium-term suicide risk, when compared to the more traditional approaches (clinician judgements, sparse logistic regression and decision trees). Similarly, Poulin et al. (2014) reported that computerised text analytics applied to unstructured medical records achieved inference accuracy of 65% or more in differentiating between veterans who died by suicide, veterans who used mental health services and did not die by suicide, and veterans who did not use mental health services nor suicided during the observation period. Analyses of language patterns of users of social networks and online communication channels has also been confirmed as a viable and promising strategy for detecting suicide-related messages and thus aiding suicide prevention efforts (Desmet & Hoste, 2018; Song et al., 2016).

A Chilean study of 707 mental health patients used a support vector machine learning approach to build a predictive model for suicide risk, which was able to recognise mental health patients vulnerable to suicide attempts or thinking about suicide with an accuracy of 78% (Barros et al., 2017). A decision tree analysis was used in a Korean study with 2754 students that participated in a mental health survey, identifying 11 sociodemographic, intra-personal, and extra-personal variables that were able to predict future suicide attempts with an accuracy of over 90% (Bae et al., 2015). Other examples of the use of computational-based methods in suicide-related research include the use of Bayesian networks (Incremental Association Markov Blankets) to investigate relationships between clinical risk factors and the repetition of suicide attempts (Lopez-Castroman et al., 2011), and the use of random forest and forward selection to determine variables associated with familial suicide attempts in a sample of suicide attempters (Gonzalez et al., 2007).

While these approaches have limitations, all the studies above report positive findings with regards to the machine learning tools used. It is increasingly apparent that data mining/machine learning approaches present cost-effective solutions over human workers, especially given the often large volume, complexity and unstructured nature of clinical data (Baca-García et al., 2006; Mukhopadhyay et al., 2013). To the authors’ knowledge, there are no published accounts available that would describe the development of a machine-learning algorithm developed specifically for the purposes of improving accuracy of identification of suicidal presentations in the emergency department administrative datasets. The importance of this work is echoed by Bernert et al. (2020) and Burke et al. (2019) who both conclude that future leveraging of the machine learning techniques will aid in the prediction and prevention of suicide.

3. The SERoSP approach and problem formulation

The SERoSP tool is part of a strategy of using machine learning methodology to augment evaluation of a leading implementation of the ZSF by GCMHSS in a large public (government funded) hospital and health service. The tool described here forms part of a wider strategy to implement machine learning to collect data which helps inform outcomes for the ZSF implementation. It should be noted that SERoSP is being used to collect ongoing data, representing a pragmatic “real-world” application of machine learning which informs a large-scale clinical implementation. For example, SERoSP has informed work on the accuracy of data on suicide-related presentations to emergency departments, showing that suicidal and self-harm presentations are under-identified in current emergency department datasets (Sveticic et al., 2020). SERoSP has been used to evaluate the GCMHSS SPS, where representations with suicide attempts to GCMHSS were analysed using time to event analysis before and after SPS implementation (Stapelberg et al., 2020) and to detail trends in suicidal presentations using 10 years of emergency department (ED) data (Stapelberg et al., 2020).

In addition to data analysis and computational optimisation techniques increasingly being applied to this clinically important field, there have been numerous calls for a standardisation of data collected on suicidal and self-harm presentations (Christensen et al., 2013; Owens et al., 2014). The World Health Organisation (World Health Organization, 2016), p. 56 states:

> In order to pursue the key objectives of a surveillance system for hospital-presented suicide attempts and self-harm, long-term sustainability is crucial. For instance, identifying suicide-attempt or self-harm patients with a risk of long-term repetition, and their characteristics, calls for obtaining data on consecutive cases of hospital-presented suicide attempts and self-harm over at least several years.

A standardised and automated tool, capable of mining complex data could contribute to such an effort, especially if adopted across a local region, such as the tool described in this paper.

Assessment of outcomes for the GCMHSS SPS, requires data on suicidal and self-harm presentations to be captured from the Emergency Department Information System (EDIS) database at the Gold Coast Hospital and Health Service (GCHHS), Australia. All consumers presenting to the two EDs of the GCHHS have data entered into EDIS on triage, including demographic data, a presenting triage text (which is entered as free text) and coded information, including a presenting complaint code, primary diagnosis International Classification of Diseases (ICD) code (OMS, 1992) and discharge destination, which is selected from dropdown menus. It is notable that no international standard exists for 'Presenting Complaint' or 'Presenting Problem' entries, which poses further challenges in analysing such presentation data (Malmström et al., 2012). The data fields captured by EDIS are shown in Table 1.

Suicidal and self-harm presentations are not uniformly coded in EDIS. Despite the existence of 24 ICD diagnostic codes which are specific to intentional self-harm presentations, ranging from X60 'Intentional self-poisoning by and exposure to non-opioid analgesics, antipyretics and antirheumatics' to X84 'Intentional self-harm by unspecified means' (OMS, 1992), coding and even triage of suicidal and self-harm presentations is heterogeneous. One challenge is that only one of the above codes, X84, is able to be coded into EDIS for self-harm presentations.

Furthermore, two Australian studies have highlighted variation in triage of mental health patients presenting to EDs. Phillips et al. (2015) showed a lack of consensus regarding dispositional outcomes at triage, suggesting a high level of subjectivity in decision-making, while Creaton et al. (2008) showed interrater concordance for triage of mental health patients in ED (using the Australasian Triage Scale) with a range of 53.3% to 65.6%. Furthermore, the distribution of triage ratings was associated with ED activity level, with a busy ED resulting in a decrease in interrater concordance (Creaton et al., 2008). In addition, presenters may not always disclose suicidality or even a suicide attempt at triage, which comes to light at a later time with a full mental health assessment. Coding by presenting complaint code and primary diagnosis ICD Code is also variable. For example, suicidal and self-harm presentation data examined here from 2015 showed that such presentations were coded using 48 different ICD primary diagnosis codes and 28 presenting complaint codes.

It could be argued that such heterogeneity necessitates human assessment of EDIS data by those engaging in data mining and coding to differentiate suicidal and self-harm presentations from other cases, e.g., accidental injuries (Kõlves et al., 2018). However, human assessment and coding of data is also generally recognised across different fields to be costly, time consuming and error prone (Glynn et al., 2012; Kieren & Munro, 1985; Stapelberg et al., 2016; Yamamoto et al., 2017). This, as well as the increasing volume and complexity of available data (Mukhopadhyay et al., 2013) has created an impetus to explore the use of sophisticated data-gathering algorithms, more recently employing machine learning or artificial intelligence frameworks.

3.1. SERoSP and the problem formulation

The aim of creating the SERoSP software program was to use machine learning to create an innovative, time and cost-effective solution for correctly identifying suicidal and self-harm presentations to EDs. The software tool thus requires high sensitivity, but also to be discriminant enough to provide as small a number of output cases as possible, i.e. avoiding false positives.

The SERoSP software tool was designed in three phases. The first phase explored the frequency of categories in various fields of an EDIS training dataset. The second phase employed an optimisation software program which used an EA with the EDIS training dataset to weight 86 variables from the data fields. The EA also weighted an additional 50 variables which represented keywords or n-grams appearing as free text in the triage field, including negations (e.g., "not suicidal") as discussed below. A total of 136 variables were weighted relative to each other with the objective function for the EA given in Eqs. (1)–(2). The algorithm outputs a SERoSP Score for each presentation by summing weights all 136 variables.

The SERoSP Score is arbitrary, as its value is based on summed weights, however the higher the SERoSP Score, the more likely the case under scrutiny is to be a suicidal presentation. A threshold for the SERoSP Score is established and if the SERoSP Score exceeds the threshold, then the case is considered a suicidal presentation. These are given in Tables B.4–B.5 for the training dataset and Tables B.6–B.7 for the validation set.[1]

$$\text{Minimise: } \sum_{i=1}^{T} |f(x_i, w) - S_i| \tag{1}$$

$$f(x_i, w) = \begin{cases} 1 & \sum_{j=1}^{n} x_{ij} w_j > P_i, \\ 0 & \text{Otherwise} \end{cases} \tag{2}$$

Where:

T is the number of cases in the training dataset,
n is the number of variables (set as 136),
w is the vector of decision variables and represents the weighting factors,
x_{ij} is the response for training case i for variable j where $x(i,j) \in \{0,1\}$,
$f(x,w)$ is the function that applies the vector w to x to determine if the presentation is case of self-harm (a value of 1), or not (a value of 0),
P_i is the SERoSP cut-off score for case i and
S_i is 1 if the presentation was self-harm, and 0 if not.

The third phase was the development of the SERoSP program, written in MATLAB (MathWorks, 2012), which employed the variables and their respective weights calculated from Phase 2, with a validation dataset to calculate sensitivity and specificity of the software tool. Fig. 1 provides a flow diagram depicting steps the process of creating training and validation datasets. The diagram details what data was used (and the dates of data capture) the number of cases used in each step and how raw data from EDIS was leveraged to progress Phase 1 (initial analysis of data), Phase 2 (classification) and Phase 3 (validation).

Phases 1 and 2 used a training dataset of EDIS data. During the months of March, April and May 2015, a total of 38,768 ED presentations occurred at the Robina and Southport campuses of Gold Coast Health, while a total of 27,374 presentations occurred in March and April 2017 (used for the validation dataset). EDIS data was obtained for this period, using the Primary Diagnostic ICD Codes and

[1] The use of a single validation dataset is consistent with the previous approach of Sveticic et al. (2020) and in other medical research (for example Dolled-Filhart et al., 2006). This psychiatrist-rated validation dataset represents a gold standard comparator and meets the clinical standard required for discrimination of suicidal case presentations.

Table 1
Emergency Department Information System (EDIS) Data Fields.

EDIS Information Fields	Explanatory Notes
Arrival Date	
Arrival Time	
URN	Unique Patient Identification Number
Name	
Age	
Sex	
ATS	Triage score based on the Australian Triage Scale
Origin	Identification of Aboriginal or Torres Strait Islander background
Diagnosis ICD Code Primary	Diagnosis coded as per ICD 10
Primary Diagnosis	Description of ICD Code
Discharge Date	
LOS	Length of stay
Status	Departure status from the emergency department
Destination	Discharge destination from the emergency department. This may be a hospital ward or discharge into the community
Presenting Complaint Code	Patient's reason for the encounter — code
Presenting Complaint Description	Patient's reason for the encounter — description
Presenting Problem	A free-text description of the presenting problem at point of triage

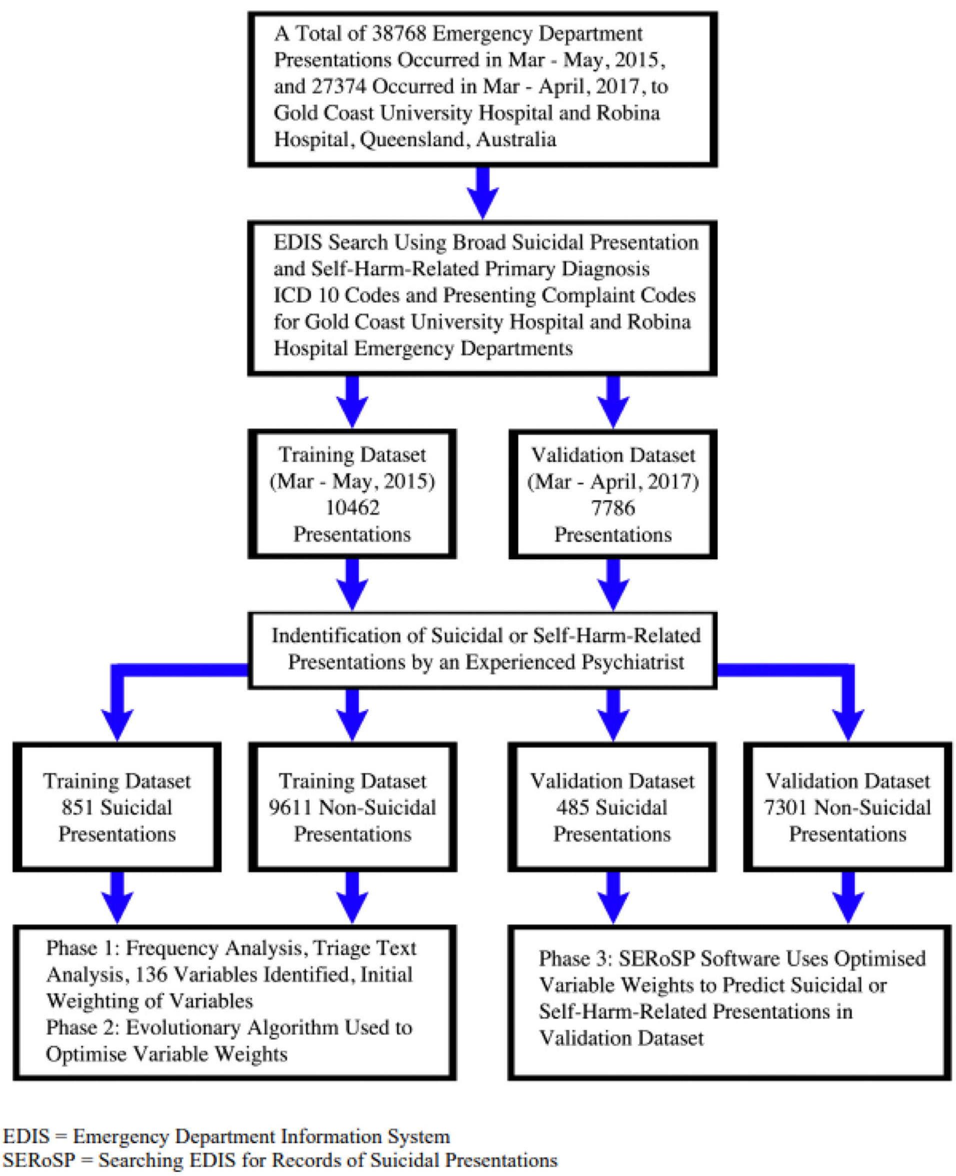

Fig. 1. An Overview of Methodology in the Development of the SERoSP Software.

Presenting Complaint Codes listed in Tables A.2 and A.3 in Appendix A respectively. This list of codes was compiled by using common codes for suicidal and self-harm presentations, but also including any codes which may include suicidal presentations, e.g., a presentation code for abdominal pain in a presenter with a paracetamol overdose. This search yielded 10,462 presentations for 2015 and 7786 for 2017 (see Fig. 1).

The EDIS entries were reviewed by a psychiatrist, yielding 851 suicidal and self-harm presentations for 2015 and 485 for 2017, which were coded in the dataset (1 = Suicidal and Self-Harm Presentation, 0 = Not a Suicidal and Self-Harm Presentation). The psychiatrist rating of presentations was based on clinical assessment of triage text and accompanying information in EDIS. Such clinical assessment of health records has been used elsewhere (Haerian et al., 2012; Kõlves et al., 2018). The rater scored conservatively, so if a case was ambiguous, it was scored 0. The rater did not attempt to separate suicide attempt presentations from non-suicidal self-injury (NSSI) or other types of suicidal or self-harm-related presentations as defined in the literature (De Leo et al., 2006; World Health Organization, 2016), as the information recorded in the EDIS is limited (Kõlves et al., 2018) and triage text and other data are entered before a full assessment of the consumer is completed.

3.2. Phase 1: Calculation of initial weights using an interactive approach

The 851 cases identified by a psychiatrist as having a suicidal and self-harm presentation were used in the calculation of the initial weights. Weighting was performed for data fields of EDIS and separately for the triage text, based on word frequency analysis with identification of key n-grams and keywords. For each field in EDIS, the different types of data elements were determined and subjected to frequency analysis to determine which elements occurred most frequently in suicidal and self-harm presentations.

Each input option for a given EDIS data field was assigned a variable. All input options for a given EDIS data field were extracted from the training dataset of 10,462 presentations. The frequency of each input option was calculated in the subset of 851 suicidal and self-harm presentations and weights were assigned to each input option accordingly. Similar methodology has been used elsewhere (Haerian et al., 2012).

An innovation based on a Human-Based Genetic Algorithm (HBGA) (Kosorukoff, 2001) approach (which is a part of the wider field of *Interactive Optimisation* (Madar et al., 2005; Parmee et al., 2001)) was used to provide initial scores for the EA. HBGAs use human agents to provide a selection of fit children for GAs (Cheng & Kosorukoff, 2004; González-Quijano et al., 2012; Kosorukoff, 2001). For example, initial input for a GA trained to perform musical phrases used an initial bank of 40 melodic excerpts recorded by a jazz pianist improvising in a similar musical context to that required by the EA (Weinberg et al., 2007). The authors state that:

> "Having a distinctly 'human' flavour, these phrases provided the GA with a rich pool of rhythmic and melodic 'genes' from which to build its own melodies. This is notably different from most standard approaches, in which the starting population is generated stochastically." (p. 353)

Similarly, initial scores for the EA in this work were chosen by a human — a psychiatrist. Initial values were chosen between −20 and 50, with an initial value chosen based on the frequency calculations, but also clinical judgement. While HBGAs may involve a fitness function based on human aesthetics, where for each generation the user determines which musical phrases remain in the population (Moroni et al., 2000; Tokui et al., 2000; Weinberg et al., 2007) over multiple generational cycles, here only the initial weightings were supplied (based on expert clinician judgement as to the relative importance of each variable). Negative weighting scores were used to weight parameters clearly linked to non-suicidal presentations, e.g., "viral gastroenteritis" in the presenting complaint.

3.3. Phase 2: The evolutionary algorithm

While many generic EA algorithms are available (genetic algorithms, particle swarm optimisation or differential evolution algorithms), this algorithm was custom-built and tailored towards its specific application. For example, initial exploratory work suggested that a high rate of mutation might be suited to the dataset. Also, innovation around choosing variables of interest (and scoring) was based on clinical expertise and also ignoring some variables to avoid bias such as age for example (see Section 5.3). SERoSP was thus purpose-built for its application, "built around the data", making it highly fit-for-purpose and arguably increasing efficiency, with the EA methodology then being applied to test the clinical assumptions and weightings made. We acknowledge that an approach of training more widely-used algorithm may also provide a good solution and we propose a comparison of SERoSP with such standard algorithms. Algorithm 1 describes the form of the EA used in this implementation of SERoSP.

Algorithm 1 The SERoSP EA. Note that the $mutation\_weight$ represents a maximum change of ±4% and was suitable for this context.

```
Read psychiatrist assigned weightings
Read presentation data
Create pop_size random solutions
Set parent specificity values to 0.85 for num_parents
for individual=1 to pop_size do
  for var=1 to solution_size do
    sol[individual][var] = Psychiatrist_start_weight[var] + 2−
    round(unif_rand(0, 1) × mutation_weight)
  end for
end for
for gen = 1 to num_gen do
  for individual = 1 to pop_size do
    Evaluate the number of true positives and true negatives using
    the presentation data
    Calculate the sensitivity for a set specificity for the individual
  end for
  Select the best num_parents individuals from sol based on lowest
  sensitivity values and assign to parents
  for parent = 1 to num_parents do
    Copy parents[parent] to sol[(parent − 1) × num_children + 1]
    for child = 2 to num_children do
      for var = 1 to solution_size do
        sol[(parent−1)]×num_children+[child][var] = parent[var]+2−
        round(unif_rand(0, 1) × mutation_weight)
      end for
    end for
  end for
end for
Output parents solution values
```

A description of each of the variables in Algorithm 1 is as follows:

- pop_size is the number of population members (set at 100 in this implementation),
- $num\_parents$ is the number of parents through which the population is formed and evolved (set at 10),
- $solution\_size$ represents the EDIS triage and field variables (136),
- $sol[x][y]$ is the decision variable and is the y^{th} weight for individual x in the population,
- $Psychiatrist\_start\_weight[x]$ is the initial weight assigned to variable x,
- $round(x)$ rounds x to the closest integer,
- $unif\_rand(x, y)$ produces a uniform random number between the bounds of x and y,
- gen is the number of generations (set to 100) and
- $parents[x]$ is parent individual x.

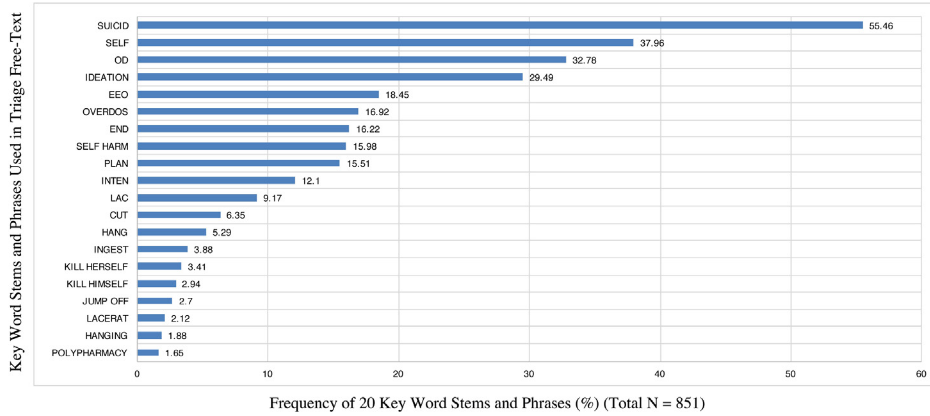

Fig. 2. Frequency of 20 keyword stems and phrases in the triage text of suicidal and self-harm presentations in the training dataset..

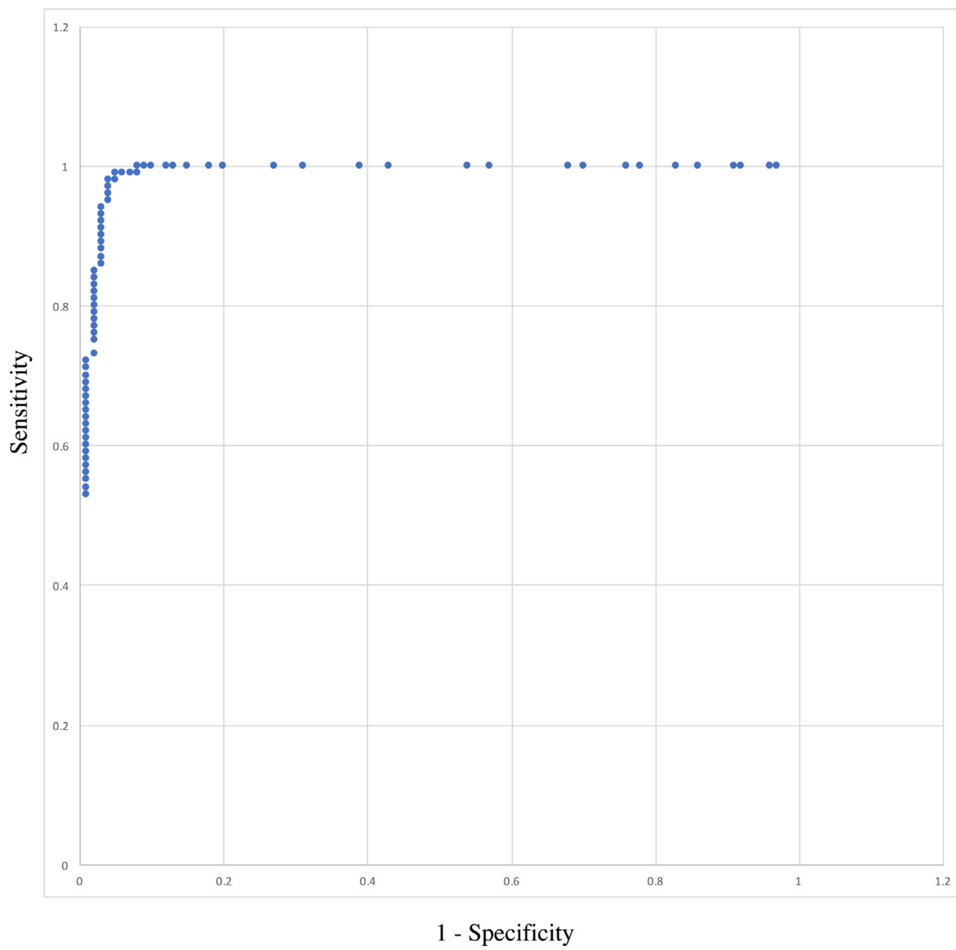

Fig. 3. Receiver Operator Curve for different SERoSP scores (2015 Training Dataset).

The current EA relies on mutation to generate new solutions rather than crossover. There have been highly effective implementations for real-world and benchmark problems that have either used a mutation only approach or high rates of mutation (Carlson-Skalak et al., 1998; Elloumi & Fortemps, 2010; Eremeev, 2018; Feeney, 2003; Fogel et al., 1991; Hughes, 2005; Shiu & Szeto, 2008; Sundaram & Venkatasubramanian, 1998; Venkatasubramanian et al., 1995). In general, these approaches can require fewer parameters and reduce issues related to premature convergence. For this application, given that 10 of 100 children in the new generation represent unmutated "best performing" children, and the other 90 are subject to mutation of all 136 weights, this gives a mutation probability of 90%, which is high. Such a high mutation probability risks introducing much random perturbation, potentially risking the potential of the algorithm to learn from the history of the search (Gen et al., 2008). However, there are applications for high mutation probability (Elloumi & Fortemps, 2010) which have been successful. Further implementations of SERoSP will test and compare different optimisation engines and parameters.

Table A.2
Primary Diagnostic ICD Codes used to search the EDIS database.

Diagnosis ICD Code Primary	Primary Diagnosis	Diagnosis ICD Code Primary	Primary Diagnosis
A08.4	Viral Intestinal Infection	S43.3	Shoulder Dislocation
B34.9	Viral Infection	S51.9	Lacerated Forearm
B37.9	Thrush	S53.40	Elbow Sprain/Strain
E10.65	Diabetes for Stabilisation	S61.9	Lacerated Hand or Wrist
E86	Dehydration	S62.6	Fracture Finger
F05.0	Delirium not superimposed on dementia, so described	S63.7	Hand Sprain/Strain
F10.0	Alcohol Intoxication	S71.0	Lacerated Hip
F10.3	Alcohol Withdrawal Syndrome	S71.1	Lacerated Thigh
F18.0	missing	S81.9	Lacerated Leg
F19.2	Drug Addiction	S83.6	Knee Sprain/Strain
F19.9	Drug Induced Mental Disorder	S91.7	Lacerated Ankle Or Foot
F20.9	Schizophrenia	T00.9	Multiple Abrasions
F29	Psychotic Episode	T01.2	Multiple Lacerations
F31.1	Bipolar Affective Disorder — Manic	T14.6	Lacerated Tendon And/Or Muscle
F31.3	Bipolar Affective Disorder — Depressed	T29.0	Burns to Multiple Areas — Unspecified
F32.3	Depression — Psychotic	T38.3	Other Antidiabetic Poisoning
F32.9	Depression	T39.0	Salicylate Poisoning
F41.0	Panic Attack	T39.1	Paracetamol Poisoning
F41.9	Anxiety	T39.3	Nonsteroidal Antiinflammatory Poisoning
F43.9	Emotional Crisis	T40.0	Opiate Toxicity
F50.0	Anorexia Nervosa	T40.5	Cocaine Poisoning
F51.0	Insomnia — Non Organic	T42.3	Poisoning by, adverse effect of and underdosing of barbiturates
F60.9	Personality Disorder	T42.4	Benzodiazepine Poisoning
F91.9	Behavioural Problems — Child	T43.5	Antipsychotic Poisoning
F99	Mental Illness — No Diagnosis	T43.68	missing
G93.1	Anoxic brain damage, not elsewhere classified	T43.69	Amphetamine Poisoning
H35.6	Retinal Haemorrhage	T43.9	Antidepressant Poisoning
H57.1	Eye — Painful	T44.3	Anticholinergic Poisoning
I20.0	Possible Cardiac Chest Pain	T44.7	Beta-Blocker Poisoning
I21.9	Myocardial Infarction — Acute	T45.0	Antihistamine Poisoning
J18.9	Pneumonia — Unspecified	T45.4	Poisoning by, adverse effect of and underdosing of iron and its compounds
J44.9	Acute Exacerbation of COPD	T45.5	Warfarin Poisoning
J90	Pleural Effusion	T46.0	Digoxin Toxicity
K82.0	Biliary Obstruction	T46.1	Poisoning by, adverse effect of and underdosing of calcium-channel blockers
K92.2	Gastrointestinal Haemorrhage	T46.41	missing
L23.9	Eczema	T50.4	Poisoning by, adverse effect of and underdosing of drugs affecting uric acid metabolism
R07.3	Non-Cardiac Chest Pain	T50.9	Other Drug Poisoning
R10.4	Abdominal Pain Recurrent	T51.8	Toxic effect of other alcohols
R11	Nausea/Vomiting — No Diagnosis	T58	Carbon Monoxide Inhalation
R27	Ataxia	T59.8	Smoke Inhalation
R45.81	Suicidal Ideation	T60.0	Organophosphate and Carbamate Poisoning
R55	Syncope/Collapse	T71	Asphyxiation, Strangulation or Hanging
S01.5	Lacerated Mouth, Lips or Oral Cavity	T75.1	Immersion
S05.0	Corneal Abrasion	T78.4	Allergic Reaction (Not Due To Serum)
S06.0	Concussion	T88.7	Other Medication Side-Effect
S11.8	Lacerated Neck	W46	Needle Stick Injury
S12.9	Fracture Cervical Spine	X84	Suicidal Ideation/Self Harm
S31.1	Lacerated Abdomen or Lower Back	Z02.7	Medical Certificate
S32.00	Fracture Lumbar Spine	Z04.8	Medical Advice on Medication
S33.7	Back Sprain/Strain	Z53.2	Did Not Wait
S41.1	Lacerated Upper Arm	Z60.9	Social Admission

4. Computational experience

This section describes the results from running SERoSP on the training and validation data. It was implemented in MATLAB and run on a standard desktop environment.

4.1. Analysis of the unstructured free-text in the presenting complaint field

In EDIS, the 'Presenting Complaint' field contains unstructured free-text, representing a triage-related clinical note, which is entered by a clinician on arrival of a patient in ED. An analysis of word frequency was carried out which aggregated all the triage text for the 851 suicidal and self-harm presentations in the training data and then generated a table of all words in the triage text and their frequency of occurrence. This yielded 2667 words used in total, from which 59 keywords were selected relevant to suicidal and self-harm presentations. The 20 most frequently occurring word stems and phrases are shown in Fig. 2. It should be noted that the percentages listed in Fig. 2 do not total 100%, as more than one key word or n-gram could occur in each triage text entry. In computational linguistics a contiguous sequence of multiple words commonly used together is referred to as an "n-gram" (Lesher et al., 1999; Martin et al., 1998). N-grams consisting of two words can be referred to as bigrams, e.g., "suicidal ideation", trigrams for three words, e.g., "loss of consciousness" and n-grams for multiple words (Lesher et al., 1999; Martin et al., 1998).

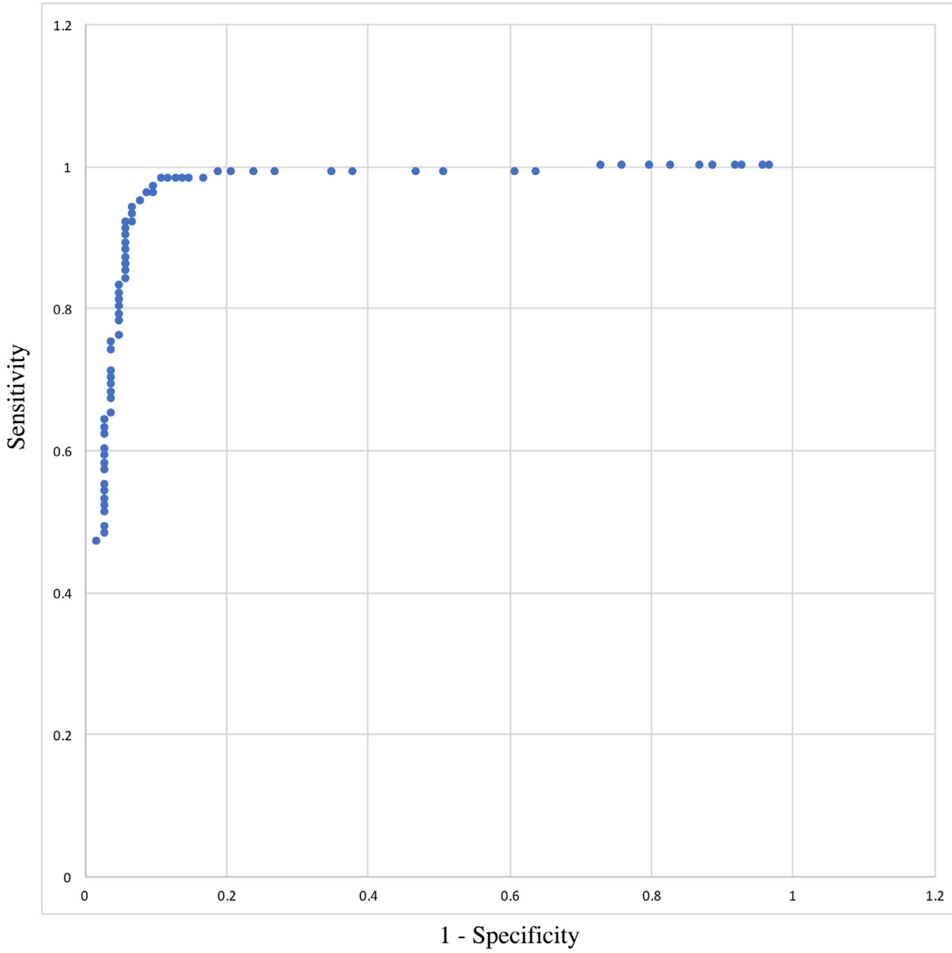

Fig. 4. Receiver Operator Curve for different SERoSP scores (2017 Validation Dataset).

Table A.3
Presenting Complaint Codes used to search the EDIS database.

Presenting Complaint Code	Presenting Complaint Description
1	Multi-Trauma
11001	Aggression
11002	Anxiety/Agitation
1400	Immersion
16003	Hallucinations
17007	Social Concern (Includes Child Protection)
21001	Altered or Loss of Consciousness
30002	Eating Disorder
30005	Suicidal–Homicidal Ideation
36001	Collapse
36018	Unsettled
8007	Swallowing Difficulty
9012	Fall
9023	Penetrating Injury
9029	Blunt Injury
9033	Laceration/Skin Tear
9034	Motorbike Crash/Quad-Driver
9038	Overdose/Toxic Exposure
9040	Strangulation/Asphyxia
9041	Crash-Other Vehicle
9042	Pedestrian Vs
9043	Suspected Airways Fb/Choking

The word frequency table was used to explore key n-grams used in the triage text entries, as well as create word stems for use with the EA (e.g., the stem "suicid" used in a search will return "suicide" and "suicidal"). Several relevant abbreviations are present in triage text, such as QAS (Queensland Ambulance Service), or QPS (Queensland Police Service), as suicidal and self-harm presenters are frequently transported to ED by police or ambulance services. Other words such as "ideation" occur frequently, which are part of longer n-grams such as "suicidal ideation". Four word stems were strongly represented in the triage text of suicidal and self-harm presentations: "Suicid*", "self*, "OD" (which is an abbreviation for "overdose") and "ideation". In addition, "EEO", which is an acronym for "Emergency Examination Order" (an instrument of the Queensland Mental Health Act 2000) is substantially represented, reflecting that several suicidal and self-harm presenters had been placed under the Mental Health Act for purposes of proceeding with a mental health examination, by ambulance or police services. The words "overdos*", "end", "self harm", "plan" and "inten*" (a stem for "intent" or "intended") were also substantially represented. A final list of 50 word stems and n-grams were formulated as variables for weighting.

4.2. The evolutionary algorithm and the training dataset

The evolutionary algorithm was first run for 50 generations with 100 children in each generation, achieving a sensitivity of 0.944 for a specificity set at a threshold of 0.95 on the training dataset. A second run of 100 generations with 100 children in each generation achieved a sensitivity of 0.992 on the training dataset at a threshold of 0.95. Manual changes were then made to the weights for each of the 136 EDIS triage and field variables: Gender bias was removed, with gender-related words such as "himself" or "herself" being weighted equally. Terms such as EEO and EEA (which are comparable legal instruments, but one is from the Queensland Mental Health Act 2000 and the latter is from the Queensland Mental Health Act 2016) were weighted equally.

Table B.4
Range sensitivities and specificities for each SERoSP cut-off score in the 2015 training dataset — Part 1.

True positives	True negatives	False positives	False negatives	Sensitivity	Specificity	SERoSP Score
851	278	9333	0	1	0.03	1
851	389	9222	0	1	0.04	2
851	746	8865	0	1	0.08	3
851	859	8752	0	1	0.09	4
851	1306	8305	0	1	0.14	5
850	1600	8011	1	1	0.17	6
850	2095	7516	1	1	0.22	7
850	2266	7345	1	1	0.24	8
849	2867	6744	2	1	0.3	9
849	3094	6517	2	1	0.32	10
849	4147	5464	2	1	0.43	11
849	4445	5166	2	1	0.46	12
849	5510	4101	2	1	0.57	13
849	5886	3725	2	1	0.61	14
849	6670	2941	2	1	0.69	15
849	7010	2601	2	1	0.73	16
849	7671	1940	2	1	0.8	17
849	7875	1736	2	1	0.82	18
849	8169	1442	2	1	0.85	19
849	8323	1288	2	1	0.87	20
848	8463	1148	3	1	0.88	21
848	8625	986	3	1	0.9	22
848	8702	909	3	1	0.91	23
847	8765	846	4	1	0.91	24
847	8825	786	4	1	0.92	25
846	8865	746	5	0.99	0.92	26
846	8920	691	5	0.99	0.93	27
846	8963	648	5	0.99	0.93	28
845	9005	606	6	0.99	0.94	29
844	9028	583	7	0.99	0.94	30
843	9068	543	8	0.99	0.94	31
840	9097	514	11	0.99	0.95	32
839	9127	484	12	0.99	0.95	33
838	9154	457	13	0.98	0.95	34
836	9179	432	15	0.98	0.96	35
834	9196	415	17	0.98	0.96	36
828	9209	402	23	0.97	0.96	37
826	9220	391	25	0.97	0.96	38
820	9235	376	31	0.96	0.96	39
816	9248	363	35	0.96	0.96	40
811	9256	355	40	0.95	0.96	41
805	9272	339	46	0.95	0.96	42
802	9281	330	49	0.94	0.97	43
797	9291	320	54	0.94	0.97	44
794	9297	314	57	0.93	0.97	45
787	9308	303	64	0.92	0.97	46
781	9315	296	70	0.92	0.97	47
777	9323	288	74	0.91	0.97	48
773	9329	282	78	0.91	0.97	49
770	9333	278	81	0.9	0.97	50
768	9338	273	83	0.9	0.97	51

Rather than lemmatising terms, words such as "self-inflicted" and "self inflicted" were given equal weighting. A Receiver Operator Characteristic (ROC) curve for the sensitivities and specificities obtained from the training dataset is shown in Fig. 3.

4.3. Validation

SERoSP was then run using the validation dataset, generating sensitivities and specificities for each cut-off score in the validation dataset. These were used to plot the ROC curve shown in Fig. 4. An optimum cutoff SERoSP Score of 36 was chosen, yielding a sensitivity of 0.95 and specificity of 0.92 for the validation dataset. While a higher SERoSP Score above the cut-off value indicates that a given case is more likely to represent a suicidal and self harm presentation, the score itself is arbitrary as stated above.

5. Discussion

After a process of frequency analysis and initial weighting, the evolutionary algorithm optimised the weighting of 136 variables to identify suicidal and self-harm presentations in EDIS data. The SERoSP program was able to discriminate suicidal from non-suicidal presentations in a validation dataset with a substantial sensitivity and specificity. The SERoSP program reliably, efficiently and cheaply identifies such presentations from EDIS data for the ongoing GCMHSS SPS evaluation. However, given the limited clinical information available on EDIS, SERoSP could be further enhanced with more accurate training data. The presentations identified by SERoSP are being examined in more detail to obtain confirmation of either 'suicide attempt', 'suicidality present', or 'self-harm', based on World Health Organisation criteria (World Health Organization, 2016), from clinical assessment records in a clinical database, the Consumer Integrated Mental Health Application (CIMHA). These more accurate assessments can then be used to reweight the EA, with the aim of creating higher sensitivity and specificity.

5.1. Overfitting and settling on local optima

Sensitivity and specificity were lower for the validation dataset than the training dataset, which can occur due to overfitting. Overfitting

Table B.5
Range sensitivities and specificities for each SERoSP cut-off score in the 2015 training dataset — Part 2.

True positives	True negatives	False positives	False negatives	Sensitivity	Specificity	SERoSP Score
766	9341	270	85	0.9	0.97	52
757	9347	264	94	0.89	0.97	53
755	9351	260	96	0.89	0.97	54
753	9357	254	98	0.88	0.97	55
745	9359	252	106	0.88	0.97	56
737	9363	248	114	0.87	0.97	57
729	9369	242	122	0.86	0.97	58
726	9371	240	125	0.85	0.98	59
722	9377	234	129	0.85	0.98	60
714	9383	228	137	0.84	0.98	61
709	9391	220	142	0.83	0.98	62
707	9396	215	144	0.83	0.98	63
703	9401	210	148	0.83	0.98	64
695	9405	206	156	0.82	0.98	65
689	9413	198	162	0.81	0.98	66
683	9418	193	168	0.8	0.98	67
671	9428	183	180	0.79	0.98	68
665	9434	177	186	0.78	0.98	69
659	9440	171	192	0.77	0.98	70
649	9445	166	202	0.76	0.98	71
642	9448	163	209	0.75	0.98	72
636	9451	160	215	0.75	0.98	73
624	9456	155	227	0.73	0.98	74
617	9459	152	234	0.73	0.98	75
611	9468	143	240	0.72	0.99	76
602	9470	141	249	0.71	0.99	77
592	9481	130	259	0.7	0.99	78
585	9485	126	266	0.69	0.99	79
579	9490	121	272	0.68	0.99	80
570	9493	118	281	0.67	0.99	81
564	9496	115	287	0.66	0.99	82
557	9501	110	294	0.65	0.99	83
551	9503	108	300	0.65	0.99	84
545	9505	106	306	0.64	0.99	85
541	9507	104	310	0.64	0.99	86
535	9511	100	316	0.63	0.99	87
526	9512	99	325	0.62	0.99	88
524	9515	96	327	0.62	0.99	89
519	9517	94	332	0.61	0.99	90
513	9519	92	338	0.6	0.99	91
512	9521	90	339	0.6	0.99	92
504	9524	87	347	0.59	0.99	93
495	9525	86	356	0.58	0.99	94
489	9530	81	362	0.57	0.99	95
483	9530	81	368	0.57	0.99	96
476	9535	76	375	0.56	0.99	97
466	9537	74	385	0.55	0.99	98
461	9538	73	390	0.54	0.99	99
454	9541	70	397	0.53	0.99	100

occurs when a machine learning algorithm models the training data too specifically, affecting its ability to generalise to other datasets (e.g., a validation dataset). This means that the algorithm (such as an EA) is weighted too much by the detail and noise in the training dataset, which may not exist in other similar datasets. When this algorithm is then run with new data, performance may be lower. Attempts to overcome overfitting were implemented here, such as checking and editing by a psychiatrist of some of the final weightings output by the EA, correcting biases such as differences in gender-specific n-grams, or the introduction of additional n-grams to the EA, not identified by text analysis in the training dataset. Despite lower performance on the validation dataset, the results were still acceptable in terms of sensitivity and specificity achieved, meeting the desired target.

A further challenge in an optimisation employing EAs is multimodal problems, where an algorithm may settle on a local optimum weighting solution, however not settling on the global optimum in the solution space (Eiben et al., 2003). The challenge of not settling on a local optimum can be addressed to some extent with a greater number of generations as well as children and it is recommended that this be attempted in future work. One solution is to employ an algorithm which has a greater random variation with lower sensitivity and specificity scores, but reduced variation for subsequent generations when children start to close in on a local solution (higher sensitivity and specificity). This would allow a greater area of the solution space to be covered initially, while allowing to focus in on local solutions when they are discovered. Strategies for enhancing search or optimisation have been widely explored, and three different approaches have been widely used (Deb et al., 2002), the first being self-adaptive evolution strategies (Bäck, 1998; Hansen & Ostermeier, 2001; Schwefel, 1988), the second being differential evolution (Storn & Price, 1997) and the third real-parameter genetic algorithms (Herrera et al., 1998).

Although it has been suggested that there are similarities in search principles between the approaches listed (Beyer & Deb, 2001; Deb et al., 2002), there is scope to improve upon the EA used here. Further improvement might also be gained using crossover rather than just mutation, of which there are several different strategies such as using two parent (Eiben et al., 2003) or multi-parent recombination (Eiben & Bäck, 1997; Tsutsui, 1998), or adopting a (1 + 1) EA model, where children must perform better than their parents from previous generations to be retained by the EA (Jansen & Wegener, 2001).

Table B.6
Range Sensitivities and Specificities for Each SERoSP Cut-Off Score in the 2017 Validation Dataset — Part 1.

True positives	True negatives	False positives	False negatives	Sensitivity	Specificity	SERoSP Score
485	230	7071	0	1	0.03	1
484	275	7026	1	1	0.04	2
484	505	6796	1	1	0.07	3
484	579	6722	1	1	0.08	4
484	776	6525	1	1	0.11	5
484	920	6381	1	1	0.13	6
484	1273	6028	1	1	0.17	7
484	1431	5870	1	1	0.2	8
484	1787	5514	1	1	0.24	9
484	1940	5361	1	1	0.27	10
482	2624	4677	3	0.99	0.36	11
482	2839	4462	3	0.99	0.39	12
482	3576	3725	3	0.99	0.49	13
482	3866	3435	3	0.99	0.53	14
482	4504	2797	3	0.99	0.62	15
482	4777	2524	3	0.99	0.65	16
482	5321	1980	3	0.99	0.73	17
480	5532	1769	5	0.99	0.76	18
480	5775	1526	5	0.99	0.79	19
479	5925	1376	6	0.99	0.81	20
477	6057	1244	8	0.98	0.83	21
477	6213	1088	8	0.98	0.85	22
477	6283	1018	8	0.98	0.86	23
476	6344	957	9	0.98	0.87	24
474	6401	900	11	0.98	0.88	25
474	6447	854	11	0.98	0.88	26
473	6483	818	12	0.98	0.89	27
473	6506	795	12	0.98	0.89	28
472	6545	756	13	0.97	0.9	29
471	6568	733	14	0.97	0.9	30
467	6593	708	18	0.96	0.9	31
467	6621	680	18	0.96	0.91	32
466	6651	650	19	0.96	0.91	33
464	6675	626	21	0.96	0.91	34
463	6698	603	22	0.95	0.92	35
461	6717	584	24	0.95	0.92	36
459	6738	563	26	0.95	0.92	37
457	6754	547	28	0.94	0.93	38
456	6764	537	29	0.94	0.93	39
455	6777	524	30	0.94	0.93	40
452	6785	516	33	0.93	0.93	41
451	6792	509	34	0.93	0.93	42
449	6802	499	36	0.93	0.93	43
448	6812	489	37	0.92	0.93	44
447	6825	476	38	0.92	0.93	45
444	6831	470	41	0.92	0.94	46
444	6838	463	41	0.92	0.94	47
440	6841	460	45	0.91	0.94	48
439	6846	455	46	0.91	0.94	49
435	6852	449	50	0.9	0.94	50
434	6857	444	51	0.89	0.94	51

5.2. Negations and text analysis used in the evolutionary algorithm

Negations were introduced as key n-grams for evaluating the triage text, with the assumption that this may assist in providing a low score for cases clearly identified on triage as not being a suicidal or self-harm presentation. The bigrams “no suicid*” and “not suicid*” were both weighted as 0 by the EA, suggesting that they were not of substantial value in discriminating suicidal from non-suicidal presentations. However, “denies suicide*” was weighted highly by the EA (Fig. 2), which was an unexpected result. Interestingly, the word “denies” appeared 136 times in triage entries of the 851 suicidal and self-harm presentations of the training dataset, while “denies suicide*” appeared 11 times. The word appeared in the context of patients documented as denying a suicide attempt after being brought in by family or emergency services with reasonable suspicion of having made a suicide attempt, with evidence of medication missing, for example.

Analysis of triage text comprised a substantial contribution to the function of the EA, with certain keywords in the triage text receiving the highest weightings (see Table B.4). Text mining approaches have proved successful in relation to identifying suicidal and self-harm behaviour (Ben-Ari & Hammond, 2015; Desmet & Hoste, 2018; Poulin et al., 2014) and with more data being available upon the migration of Gold Coast Health to the FirstNet database (Cerner, North Kansas City, Missouri), text mining approaches could be expanded. Furthermore, the text analysis employed here could be expanded upon with more sophisticated semantic and content analysis methodology, for example network analysis (Ignatow, 2016; Roberts & Popping, 1996) could be employed to understand what groups of words cluster together in triage text for suicidal and self-harm versus non-suicidal and self-harm presentations. It is further recommended that the EA be retrained using SNOMED codes (Cornet & de Keizer, 2008; Cote & Robboy, 1980) upon system migration to FirstNet.

5.3. Exclusion of variables, such as age

Most, but not all, EDIS data fields were used in designing the EA used here. One example of a field not used is age of presenters. which could have been used as a parameter and weighted as part of the

Table B.7
Range Sensitivities and Specificities for Each SERoSP Cut-Off Score in the 2017 Validation Dataset — Part 2.

True positives	True negatives	False positives	False negatives	Sensitivity	Specificity	SERoSP Score
430	6865	436	55	0.89	0.94	52
426	6870	431	59	0.88	0.94	53
422	6873	428	63	0.87	0.94	54
416	6881	420	69	0.86	0.94	55
414	6881	420	71	0.85	0.94	56
412	6887	414	73	0.85	0.94	57
407	6895	406	78	0.84	0.94	58
404	6906	395	81	0.83	0.95	59
402	6912	389	83	0.83	0.95	60
399	6921	380	86	0.82	0.95	61
392	6925	376	93	0.81	0.95	62
386	6933	368	99	0.8	0.95	63
383	6941	360	102	0.79	0.95	64
378	6953	348	107	0.78	0.95	65
367	6961	340	118	0.76	0.95	66
364	6977	324	121	0.75	0.96	67
357	6986	315	128	0.74	0.96	68
345	6992	309	140	0.71	0.96	69
344	6999	302	141	0.71	0.96	70
342	7005	296	143	0.71	0.96	71
339	7011	290	146	0.7	0.96	72
336	7016	285	149	0.69	0.96	73
331	7021	280	154	0.68	0.96	74
324	7027	274	161	0.67	0.96	75
317	7034	267	168	0.65	0.96	76
314	7037	264	171	0.65	0.96	77
313	7042	259	172	0.65	0.96	78
308	7049	252	177	0.64	0.97	79
304	7053	248	181	0.63	0.97	80
299	7056	245	186	0.62	0.97	81
293	7058	243	192	0.6	0.97	82
290	7061	240	195	0.6	0.97	83
288	7065	236	197	0.59	0.97	84
286	7068	233	199	0.59	0.97	85
282	7074	227	203	0.58	0.97	86
277	7078	223	208	0.57	0.97	87
269	7083	218	216	0.55	0.97	88
262	7084	217	223	0.54	0.97	89
261	7087	214	224	0.54	0.97	90
260	7091	210	225	0.54	0.97	91
256	7096	205	229	0.53	0.97	92
254	7096	205	231	0.52	0.97	93
254	7100	201	231	0.52	0.97	94
251	7102	199	234	0.52	0.97	95
245	7107	194	240	0.51	0.97	96
240	7110	191	245	0.49	0.97	97
235	7111	190	250	0.48	0.97	98
232	7116	185	253	0.48	0.97	99
227	7121	180	258	0.47	0.98	100

algorithm, however was also recognised as potentially introducing unwanted bias. The number of children and early adolescents presenting with suicidal and self-harm presentations occurs with less frequency than adult presentations. However, while suicide accounts for 1.9% of total mortality in Australia overall (Australian Bureau of Statistics, 2016), suicide accounts for a much higher proportion of deaths among younger Australians, with over one-third of deaths (35.9%) among people 15–24 years of age, and over a quarter of deaths (28.6%) among those 25–34 years of age being due to suicide (Australian Bureau of Statistics, 2016).

Suicide was the leading cause of death among all people 15–44 years of age (Australian Bureau of Statistics, 2016), emphasising the importance of younger presenters and making their detection paramount. Given their relatively lower frequency of presentation, children or adolescents might possibly receive a lower score in the SERoSP software, if age had been taken into account in the initial weighting process. Thus a decision was made not to include this parameter.

6. Conclusions

The SERoSP program is a reliable and cost-effective tool for identification of suicidal and self-harm presentations from EDIS data and is currently being used in the GCMHSS SPS evaluation, demonstrating its utility. Future work should include retraining the EA with larger datasets, spanning larger periods of time. If retraining of the EA is undertaken, presentation dates can be used in addition to the other parameters. Presentation patterns over day of the week, time of month or presentation patterns across the year might provide additional useful data which can be leveraged by the EA. It is also recommended that training and validation datasets could benefit from being checked against other sources of data. Checking against clinical databases which contain psychiatric assessment information, such as CIMHA to verify suicidal and self-harm presentations is recommended. The pursuit of data linkage by health systems (for example, Bates et al., 2018) could enable such endeavours.

Ongoing work aims to address some of the current limitations of SERoSP and improve its efficiency and efficacy. One way that this can be achieved is by implementing it using different search algorithms, such as differential evolution or particle swarm optimisation. As the task is primarily one of classification, traditional classification systems like Support Vector Machines or K-nearest neighbour, will be tested and compared against the current results of the initial algorithm. Another aspect of the work is the expansion of datasets that are currently being used, as they currently only reflect one region's data. In future, SERoSP could tested to see if it could be used more generally to data mine other mental health diagnostic groups, such as presentations with psychotic illness, substance use presentations or people presenting with mood disturbance.

CRediT authorship contribution statement

Nicolas J.C. Stapelberg: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Marcus Randall:** Methodology, Writing - original draft, Writing - review & editing. **Jerneja Sveticic:** Methodology, Data curation, Resources, Writing - review & editing. **Pete Fugelli:** Methodology, Data curation, Resources, Writing - review & editing. **Hasmeera Dave:** Resources (lit review), Writing - original draft, Writing - review & editing. **Kathryn Turner:** Conceptualization, Resources, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. ICD codes

See Tables A.2 and A.3.

Appendix B. Range sensitivities and SERoSP cut-off score

See Tables B.4–B.7.

References

Australian Bureau of Statistics (2016). 3303.0 - Causes of death, Australia. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/bySubject/3303.0~2017~MainFeatures~Intentionalself-harm~key-characteristics~3.

Baca-García, E., Perez-Rodriguez, M. M., Basurte-Villamor, I., Saiz-Ruiz, J., Leiva-Murillo, J. M., de Prado-Cumplido, M., Santiago-Mozos, R., Artés-Rodríguez, A., & De Leon, J. (2006). Using data mining to explore complex clinical decisions: A study of hospitalization after a suicide attempt. *Journal of Clinical Psychiatry*, *67*(7), 1124–1132.

Bäck, T. (1998). An overview of parameter control methods by self-adaptation in evolutionary algorithms. *Fundamenta Informaticae*, *35*(1–4), 51–66.

Bae, S. M., Lee, S. A., & Lee, S.-H. (2015). Prediction by data mining, of suicide attempts in Korean adolescents: a national study. *Neuropsychiatric Disease and Treatment*, *11*, 2367.

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. *Brazilian Journal of Psychiatry*, *39*(1), 1–11.

Bates, D., Heitmueller, A., Kakad, M., & Saria, S. (2018). Why policymakers should care about "big data" in healthcare. *Health Policy and Technology*, *7*(2), 211–216.

Ben-Ari, A., & Hammond, K. (2015). Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian Gulf War. In *48th Hawaii international conference on system sciences* (pp. 3168–3175). IEEE.

Bernert, R., Hilberg, A., Melia, R., Kim, J., Shah, N., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: A systematic review of machine learning investigations 17. (p. 5929).

Beyer, H.-G., & Deb, K. (2001). On self-adaptive features in real-parameter evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, *5*(3), 250–270.

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of Affective Disorders*, *245*, 869–884.

Carlson-Skalak, S., White, M. D., & Teng, Y. (1998). Using an evolutionary algorithm for catalog design. *Research in Engineering Design*, *10*(2), 63–83.

Carmona, C. J., González, P., Del Jesus, M., Navío-Acosta, M., & Jiménez-Trevino, L. (2011). Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department. *Soft Computing*, *15*(12), 2435–2448.

Cheng, C. D., & Kosorukoff, A. (2004). Interactive one-max problem allows to compare the performance of interactive and human-based genetic algorithms. In *Genetic and evolutionary computation conference* (pp. 983–993). Springer.

Christensen, H., Beautrais, A., Shand, F., AM, M. D. O., Petrie, M. K., & Buckley, M. H. (2013). Literature review for the development of the report card.

Coffey, M. J., & Coffey, C. E. (2016). How we dramatically reduced suicide. *NEJM Catalyst*.

Cornet, R., & de Keizer, N. (2008). Forty years of SNOMED: A literature review. In *BMC medical informatics and decision making (vol. 8)* (p. S2). BioMed Central.

Cote, R. A., & Robboy, S. (1980). Progress in medical information management: Systematized nomenclature of medicine (SNOMED). *Jama*, *243*(8), 756–762.

Covington, D., Hogan, M., Abreu, J., Berman, A., Breux, P., Coffey, E., Comeau, C., Comtois, K., Damle, C., & Davidson, L. (2011). Suicide care in systems framework. *National Action Alliance: Clinical Care and Intervention Task Force*.

Creaton, A., Liew, D., Knott, J., & Wright, M. (2008). Interrater reliability of the Australasian triage scale for mental health patients. *Emergency Medicine Australasia*, *20*(6), 468–474.

De Leo, D., Burgis, S., Bertolote, J. M., Kerkhof, A. J., & Bille-Brahe, U. (2006). Definitions of suicidal behavior: Lessons learned from the WHO/EURO Multicentre study. *Crisis*, *27*(1), 4–15.

Deb, K., Anand, A., & Joshi, D. (2002). A computationally efficient evolutionary algorithm for real-parameter optimization. *Evolutionary Computation*, *10*(4), 371–395.

Desmet, B., & Hoste, V. (2018). Online suicide prevention through optimised text classification. *Information Sciences*, *439*, 61–78.

Dolled-Filhart, M., Rydén, L., Cregger, M., Jirström, K., Harigopal, M., Camp, R. L., & Rimm, D. L. (2006). Classification of breast cancer using genetic algorithms and tissue microarrays. *Clinical Cancer Research*, *12*(21), 6459–6468.

Eiben, A. E., & Bäck, T. (1997). Empirical investigation of multiparent recombination operators in evolution strategies. *Evolutionary Computation*, *5*(3), 347–365.

Eiben, A. E., & Smith, J. E. (2003). *Introduction to evolutionary computing (vol. 53)*. Springer.

Elloumi, S., & Fortemps, P. (2010). A hybrid rank-based evolutionary algorithm applied to multi-mode resource-constrained project scheduling problem. *European Journal of Operational Research*, *205*(1), 31–41.

Eremeev, A. V. (2018). On proportions of fit individuals in population of mutation-based evolutionary algorithm with tournament selection. *Evolutionary Computation*, *26*(2), 269–297.

Feeney, B. (2003). *Determining optimum and near-optimum golomb rulers using genetic algorithms* (Master's thesis), University College Cork.

Fogel, D. B., Fogel, L. J., & Atmar, J. (1991). Meta-evolutionary programming. In *Conference record of the twenty-fifth asilomar conference on signals, systems & computers* (pp. 540–541). IEEE Computer Society.

Gen, M., Cheng, R., & Lin, L. (2008). *Network models and optimization: Multiobjective genetic algorithm approach*. Springer Science & Business Media.

Glynn, L. H., Hallgren, K. A., Houck, J. M., & Moyers, T. B. (2012). CACTI: Free, open-source software for the sequential coding of behavioral interactions. *PLoS One*, *7*(7), Article e39740.

Gonzalez, M., Basurte-Villamor, I., Saiz-Ruiz, J., & Leiva-Murillo, J. M. (2007). Variables associated with familial suicide attempts in a sample of suicide attempters. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *31*(6), 1312–1316.

González-Quijano, J., Abderrahim, M., Bensalah, C., & Al-kaff, A. (2012). A human-based genetic algorithm applied to the problem of learning in-hand manipulation tasks. In *Proceedings of the international conference on intelligent robots and systems (vol. 2012)*.

Goodwin, L., VanDyne, M., Lin, S., & Talbert, S. (2003). Data mining issues and opportunities for building nursing knowledge. *Journal of Biomedical Informatics*, *36*(4–5), 379–388.

Haerian, K., Salmasian, H., & Friedman, C. (2012). Methods for identifying suicide or suicidal ideation in EHRs. In *AMIA annual symposium proceedings (vol. 2012)* (p. 1244). American Medical Informatics Association.

Hansen, N., & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, *9*(2), 159–195.

Herrera, F., Lozano, M., & Verdegay, J. L. (1998). Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, *12*(4), 265–319.

Hughes, E. J. (2005). Checkers using a co-evolutionary on-line evolutionary algorithm. In *IEEE congress on evolutionary computation (vol. 2)* (pp. 1899–1905). IEEE.

Ignatow, G. (2016). Theoretical foundations for digital text analysis. *Journal for the Theory of Social Behaviour*, *46*(1), 104–120.

Jansen, T., & Wegener, I. (2001). Evolutionary algorithms-how to cope with plateaus of constant fitness and when to reject strings of the same fitness. *IEEE Transactions on Evolutionary Computation*, *5*(6), 589–599.

Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2019). Suicidal ideation detection: A review of machine learning methods and applications. arXiv preprint arXiv:1910.12611.

Joubert, L., Petrakis, M., & Cementon, E. (2012). Suicide attempt presentations at the emergency department: Outcomes from a pilot study examining precipitating factors in deliberate self-harm and issues in primary care physician management. *Social Work in Health Care*, *51*(1), 66–76.

Kieren, D., & Munro, B. (1985). *The observational recording dilemma*: *(Technical report no. 143)*, Edmonton, AB: ERIC, University of Alberta.

Kõlves, K., Crompton, D., Turner, K., Stapelberg, N. J., Khan, A., Robinson, G., & De Leo, D. (2018). Trends and repetition of non-fatal suicidal behaviour: Analyses of the Gold Coast University Hospital's Emergency Department. *Australasian Psychiatry*, *26*(2), 170–175.

Kosorukoff, A. (2001). Human based genetic algorithm. In *IEEE international conference on systems, man and cybernetics. E-systems and E-man for cybernetics in cyberspace cat. no. 01CH37236 (vol. 5)*. (pp. 3464–3469).

Lesher, G. W., Moulton, B. J., & Higginbotham, D. J. (1999). Effects of ngram order and training text size on word prediction. In *Proceedings of the RESNA'99 annual conference* (pp. 52–54). Citeseer.

Lopez-Castroman, J., de las Mercedes Perez-Rodriguez, M., Jaussent, I., Alegria, A. A., Artes-Rodriguez, A., Freed, P., Guillaume, S., Jollant, F., Leiva-Murillo, J. M., & Malafosse, A. (2011). Distinguishing the relevant features of frequent suicide attempters. *Journal of Psychiatric Research*, *45*(5), 619–625.

Madar, J., Abonyi, J., & Szeifert, F. (2005). Interactive particle swarm optimization. In *5th International conference on intelligent systems design and applications* (pp. 314–319). IEEE.

Malmström, T., Huuskonen, O., Torkki, P., & Malmström, R. (2012). Structured classification for ED presenting complaints–from free text field-based approach to ICPC-application. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *20*(1), 76.

Martin, S., Liermann, J., & Ney, H. (1998). Algorithms for bigram and trigram word clustering. *Speech Communication*, *24*(1), 19–37.

MathWorks, I. (2012). *MATLAB and statistics toolbox release 2012b*. Natick (Massachusetts, United States).

Mersey Care (2015). Sd38 Zero suicide policy. www.merseycare.nhs.uk.

Moroni, A., Manzolli, J., Zuben, F. V., & Gudwin, R. (2000). Vox populi: An interactive evolutionary system for algorithmic music composition. *Leonardo Music Journal*, 49–54.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2013). A survey of multiobjective evolutionary algorithms for data mining: Part I. *IEEE Transactions on Evolutionary Computation*, *18*(1), 4–19.

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

Office of the Surgeon General (US and National Action Alliance for Suicide Prevention) (2012). *2012 national strategy for suicide prevention: goals and objectives for action: A report of the US surgeon general and of the national action alliance for suicide prevention*. US Department of Health & Human Services (US).

OMS (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva.

Oquendo, M., Baca-Garcia, E., Artes-Rodriguez, A., Perez-Cruz, F., Galfalvy, H., Blasco-Fontecilla, H., Madigan, D., & Duan, N. (2012). Machine learning and data mining: Strategies for hypothesis generation. *Molecular Psychiatry*, *17*(10), 956.

Owens, C., Roberts, S., & Taylor, J. (2014). Utility of local suicide data for informing local and national suicide prevention strategies. *Public Health*, *128*(5), 424–429.

Parmee, I., Cvetkovic, D., Bonham, C., & Packham, I. (2001). Introducing prototype interactive evolutionary systems for ill-defined, multi-objective design environments. *Advances in Engineering Software*, *32*(6), 429–441.

Phillips, G., Gerdtz, M. F., Elsom, S. J., Weiland, T. J., & Castle, D. (2015). Mental health nurses' dispositional decision-making for people presenting to the emergency department with deliberate self-harm: An exploratory study. *Perspectives in Psychiatric Care*, *51*(2), 148–153.

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), Article e85733.

Roberts, C. W., & Popping, R. (1996). Themes, syntax and other necessary steps in the network analysis of texts: A research paper. *Social Science Information*, *35*(4), 657–665.

Salinas-Pérez, J. A., Rodero-Cosano, M. L., García-Alonso, C. R., & Salvador-Carulla, L. (2015). Applying an evolutionary algorithm for the analysis of mental disorders in macro-urban areas: The case of barcelona. *Spatial Economic Analysis*, *10*(3), 270–288.

Schwefel, H.-P. (1988). Collective intelligence in evolving systems. In *Ecodynamics* (pp. 95–100). Springer.

Shiu, K. L., & Szeto, K. (2008). Self-adaptive mutation only genetic algorithm: An application on the optimization of airport capacity utilization. In C. Fyfe, D. Kim, S.-Y. Lee, & H. Yin (Eds.), *Intelligent data engineering and automated learning* (pp. 428–435). Springer Berlin Heidelberg.

Song, J., Song, T. M., Seo, D.-C., & Jin, J. H. (2016). Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. *Journal of Adolescent Health*, *59*(6), 668–673.

Stapelberg, N. J., Neumann, D. L., Shum, D. H., Mcconnell, H., & Hamilton-Craig, I. (2016). A preprocessing tool for removing artifact from cardiac RR interval recordings using three-dimensional spatial distribution mapping. *Psychophysiology*, *53*(4), 482–492.

Stapelberg, N. J., Sveticic, J., Hughes, I., Almeida-Crasto, A., Gaee-Atefi, T., Gill, N., & ...Van Engelen, H. (2020). Efficacy of the Zero Suicide framework in reducing recurrent suicide attempts: cross-sectional and time-to-recurrent-event analyses. *British Journal of Psychiatry*, 1–10.

Stapelberg, N. J., Sveticic, J., Hughes, I., & Turner, K. (2020). Suicidal presentations to emergency departments in a large Australian public health service over 10 years. *International journal of environmental research and public health*, *17*(16), 5920.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359.

Sundaram, A., & Venkatasubramanian, V. (1998). Parametric sensitivity and search-space characterization studies of genetic algorithms for computer-aided polymer design. *Journal of Chemical Information and Computer Sciences*, *38*(6), 1177–1191.

Sveticic, J., Stapelberg, N. C., & Turner, K. (2020). Suicidal and self-harm presentations to emergency departments: The challenges of identification through diagnostic codes and presenting complaints. *Health Information Management Journal*, *49*(1), 38–46.

Tokui, N., & Iba, H. (2000). Music composition with interactive evolutionary computation. In *Proceedings of the third international conference on generative art (vol. 17)*. (pp. 215–226).

Tsutsui, S. (1998). Multi-parent recombination in genetic algorithms with search space boundary extension by mirroring. In *International conference on parallel problem solving from nature* (pp. 428–437). Springer.

Venkatasubramanian, V., Chan, K., & Caruthers, J. M. (1995). Evolutionary design of molecules with desired properties using the genetic algorithm. *Journal of Chemical Information and Computer Sciences*, *35*(2), 188–195.

Weinberg, G., Godfrey, M., Rae, A., & Rhoads, J. (2007). A real-time genetic algorithm in human-robot musical improvisation. In *International symposium on computer music modeling and retrieval* (pp. 351–359). Springer.

World Health Organization (2016). Practice manual for establishing and maintaining surveillance systems for suicide attempts and self-harm.

World Health Organization (2017). Mental health. http://www.who.int/mental_health/suicide-prevention/en/.

Yamamoto, K., He, Q., Shin, H. J., & von Davier, M. (2017). Developing a machine-supported coding system for constructed-response items in PISA. *ETS Research Report Series*, *2017*(1), 1–15.