



Identification of Suicidal Ideation in the Canadian Community Health Survey—Mental Health Component Using Deep Learning

Sneha Desai^{1,2†}, Myriam Tanguay-Sela^{2,3†}, David Benrimoh^{2,4,5*}, Robert Fratila², Eleanor Brown^{2,6}, Kelly Perlman^{2,7}, Ann John⁸, Marcos DelPozo-Banos⁸, Nancy Low⁴, Sonia Israel², Lisa Palladini^{4,5} and Gustavo Turecki^{4,7}

OPEN ACCESS

Edited by:

Bertram Müller-Myhsok,
Max Planck Institute of Psychiatry
(MPI), Germany

Reviewed by:

Shailesh Tripathi,
Tampere University of Technology,
Finland
Marcus Ising,
Max Planck Institute of Psychiatry
(MPI), Germany

*Correspondence:

David Benrimoh
david.benrimoh@mail.mcgill.ca

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Medicine and Public Health,
a section of the journal
Frontiers in Artificial Intelligence

Received: 12 May 2020

Accepted: 25 May 2021

Published: 24 June 2021

Citation:

Desai S, Tanguay-Sela M, Benrimoh D, Fratila R, Brown E, Perlman K, John A, DelPozo-Banos M, Low N, Israel S, Palladini L and Turecki G (2021) Identification of Suicidal Ideation in the Canadian Community Health Survey—Mental Health Component Using Deep Learning. *Front. Artif. Intell.* 4:561528. doi: 10.3389/frai.2021.561528

¹Department of Computer Science, University of Toronto, Toronto, ON, Canada, ²Aifred Health Inc., Montreal, QC, Canada, ³Montreal Neurological Institute, McGill University, Montreal, QC, Canada, ⁴Department of Psychiatry, McGill University, Montreal, QC, Canada, ⁵Faculty of Medicine, McGill University, Montreal, QC, Canada, ⁶Department of Psychological and Brain Sciences, Boston University, Boston, MA, United States, ⁷Douglas Mental Health University Institute, Montreal, QC, Canada, ⁸Swansea University, Swansea, United Kingdom

Introduction: Suicidal ideation (SI) is prevalent in the general population, and is a risk factor for suicide. Predicting which patients are likely to have SI remains challenging. Deep Learning (DL) may be a useful tool in this context, as it can be used to find patterns in complex, heterogeneous, and incomplete datasets. An automated screening system for SI could help prompt clinicians to be more attentive to patients at risk for suicide.

Methods: Using the Canadian Community Health Survey—Mental Health Component, we trained a DL model based on 23,859 survey responses to classify patients with and without SI. Models were created to classify both lifetime SI and SI over the last 12 months. From 582 possible parameters we produced 96- and 21-feature versions of the models. Models were trained using an undersampling procedure that balanced the training set between SI and non-SI; validation was done on held-out data.

Results: For lifetime SI, the 96 feature model had an Area under the receiver operating curve (AUC) of 0.79 and the 21 feature model had an AUC of 0.77. For SI in the last 12 months the 96 feature model had an AUC of 0.71 and the 21 feature model had an AUC of 0.68. In addition, sensitivity analyses demonstrated feature relationships in line with existing literature.

Discussion: Although further study is required to ensure clinical relevance and sample generalizability, this study is an initial proof of concept for the use of DL to improve identification of SI. Sensitivity analyses can help improve the interpretability of DL models. This kind of model would help start conversations with patients which could lead to improved care and a reduction in suicidal behavior.

Keywords: deep learning, suicidal ideation, risk assessment, predictors, machine learning, artificial intelligence, Canadian community health survey, Canadian community health survey—mental health 2012

INTRODUCTION

Suicide is one of the leading causes of death across the world, accounting for approximately 800,000 deaths each year, with the number of attempts an order of magnitude higher [World Health Organization (WHO), 2018]. Globally, suicide accounts for 16% of injury deaths [World Health Organization (WHO), 2012] and is the second leading cause of death in young people aged 15–29 years [World Health Organization (WHO), 2014]. This makes suicide prevention a major public health concern (Turecki and Brent, 2016). According to a meta-analysis of 365 studies, among the most important risk factors for suicide attempts and deaths are previous self-injurious behaviors and suicidal ideation (Franklin et al., 2017). Suicidal ideation includes any thoughts about suicide such as a desire for or planning of a suicide attempt and must be distinguished from actual suicidal attempts which involve acting on these thoughts (Beck et al., 1979). This is addressed by item 9 of the depression module of the Patient Health Questionnaire (PHQ-9) as “thoughts that you would be better off dead or of hurting yourself in some way” (Kroenke and Spitzer, 2002). Importantly, there is a moderately strong association between suicide and suicidal ideation, making it an important factor to consider when assessing suicide risk (Hubers et al., 2016; McHugh et al., 2019). It is important to note that this association is heterogeneous and has low positive predictive value and sensitivity (McHugh et al., 2019). As such, it is clear that not all individuals who die by suicide will have previously expressed suicidal ideation. On the other hand, suicidal ideation is much more common than attempts, and many patients who express suicidal ideation do not actually attempt suicide (Srivastava and Kumar, 2005). Regardless, proactive detection of ideation is helpful in the identification of patients at risk of suicide.

In current clinical practice, the primary method for identifying the presence of suicidal ideation is through direct questioning or patient self-report. Suicidal ideation can also be identified and characterized using instruments, such as the PHQ-9 (Kroenke et al., 2001). This method is limited because patients may conceal suicidal intentions from clinicians, who often fail to even ask about suicidal ideation (Bongiovi-Garcia et al., 2009). It would therefore be clinically useful to identify which patients may be at risk of suicidal ideation without needing to ask them directly, perhaps by using an automated screening system incorporated into the electronic medical record, as this would allow clinicians to identify patients who might benefit from further assessment and resources.

In the current literature, the vast majority of studies focus on identifying individual predictors or an interaction of only a few factors, resulting in small effect sizes with low predictive value (Franklin et al., 2017). As such, it may be useful to employ more sophisticated methods that can consider a large number of factors when making classifications. Machine learning, which allows for the creation of models that can consider many factors and identify complex relationships between them, may be an ideal tool for identifying people with suicidal ideation. While a few machine learning models have been created to predict suicide attempts or suicidal behavior (Barak-Corren et al., 2016; Passos et al., 2016; Walsh et al., 2017; DelPozo-Banos et al., 2018), we

found only two that aimed at predicting suicidal ideation: Jordan et al. (2018), which focused on predicting suicidal ideation in primary care using PHQ-9 items, and Ryu et al. (2018) which demonstrated a suicidal ideation prediction model on a matched SI/non-SI sample derived from general population data. To our knowledge, we are the first group to create a model for suicidal ideation in a realistic general population, and to determine if sensitivity analyses could assist in improving the interpretability of this model.

Our objective was to train a model to identify suicidal ideation in the general population in order to include potential suicide victims who would not seek medical attention prior to their suicide attempt or who have infrequent contact with clinicians. With this goal in mind, we chose to use a deep learning model for a number of reasons. Firstly, deep learning models can be robust to missing data (Cai et al., 2018), which is common in clinical datasets. More importantly, these models are designed to find complex, non-linear patterns in data without requiring the user to specify mediators or moderators, allowing for a better approximation of the intricate relationships between the multitude of variables that put an individual at risk for suicidal thoughts.

Ideally, our model would be paired with a clinical decision support system (CDSS) that alerts clinicians and other healthcare practitioners to patients who may require further assessment and monitoring of possible suicidal thoughts. Such a tool would connect patients with their clinicians, allowing patients to fill out requested questionnaires and track their progress, while providing clinicians with an organized interface to follow the profiles of their individual patients. Similar tools have been found to be clinically useful in detecting and reducing sepsis mortality, and predicting oral cancer recurrence (Exarchos et al., 2012; Manaktala and Claypool, 2017).

Additionally, we hoped to use our machine learning approach to elucidate which patient characteristics are involved in determining the risk for suicidal ideation. This is important from a clinical perspective for understanding the factors that might cause suicidal ideation in an individual person. It is also valuable from a public health perspective, as we may discover risk factors for suicidal ideation amenable to intervention via social programs. Our approach is novel in that our sensitivity analysis allows for the identification of potential risk and protective factors.

MATERIALS AND METHODS

Dataset

The Canadian Community Health Survey—Mental Health Component is a publicly available database provided by Statistics Canada. We therefore had no direct interactions with any patients for our study. We received ethics approval to analyze the data from the Douglas Mental Health University Institute Research Ethics Board (IUSMD-17-39). Data was collected in 2012 cross-sectionally for 25,113 people of ages 15 and over living in the ten provinces of Canada. The data was collected either by telephone or in person and 582 data points were collected per respondent. Participants were asked whether they had experienced suicidal ideation in their lifetime and in the last 12 months. The

BOX 1 | Reproduced here from the CCHS interview guide are the specific questions asked by the interviewer about lifetime and last-12-months suicidal ideation:

- LIFETIME
Has EXPERIENCE A (You seriously thought about committing suicide or taking your own life) ever happened to you?
- LAST 12 MONTHS
In the past 12 months, did EXPERIENCE A (You seriously thought about committing suicide or taking your own life) happen to you?

questions about SI were asked during an interview; see **Box 1** for details. The response options were “yes”, “no”, “not applicable”, “don’t know”, “refusal” and “not stated”. As such, no information was provided regarding the frequency or precise timing of the suicidal ideation. We attempted to correctly identify participant answers to the lifetime and last 12 months questions separately. We included only subjects who gave a firm “yes” or “no” to the questions about suicidal ideation to maximize the discriminative ability of our model. This reduced our sample size for the identification of lifetime suicidal ideation to 23,859 with 21,597 responding “no” and 2,262 responding “yes” and the sample size for the 12 months suicidal ideation identification to 3,441 with 2,512 responding “no” and 929 responding “yes”. The size and makeup of both these subsets of the data are summarized in **Supplementary Table S6**. There were 485 people who responded “yes” to both questions.

Models

The neural network used was a feed-forward fully-connected network with three hidden layers of 400 neurons each activated by the scaled exponential linear unit (SELU) function (Klambauer et al., 2017). SELU activation paired with AlphaDropout at a 50% dropout rate for each of the hidden layers (Klambauer et al., 2017) maintains a self-normalizing property of the trained parameters of the network so as to keep the training procedure stable. Adam optimization (Kingma and Ba, 2015) with a learning rate of $1e^{-4}$ was used to train the network’s loss function configured as categorical cross entropy. The final prediction layer had a softmax activation, allowing the network to establish its prediction in the form of a probability for both output classes. Different neural network architectures were tried and results are presented in **Supplementary Figure S3** of our **Supplementary Materials**. Performance decrease is apparent when the network has less than two layers or more than three layers. Thus, we chose a middle ground of three hidden layers of 400 neurons each for the rest of our analysis.

As baseline models for comparison, we include results from random forest and gradient boosting models. The random forest classifier (Breiman, 2001) was configured with 100 estimators (i.e., composed of 100 decision trees) and used the Gini impurity entropy calculation to determine the decision boundaries. It was implemented from the Scikit-learn Python package (Pedregosa et al., 2011) using all the default parameters except for those specified earlier. The gradient boosting classifier was configured with the default configurations of the Scikit-learn Python package (Pedregosa et al., 2011).

Approach

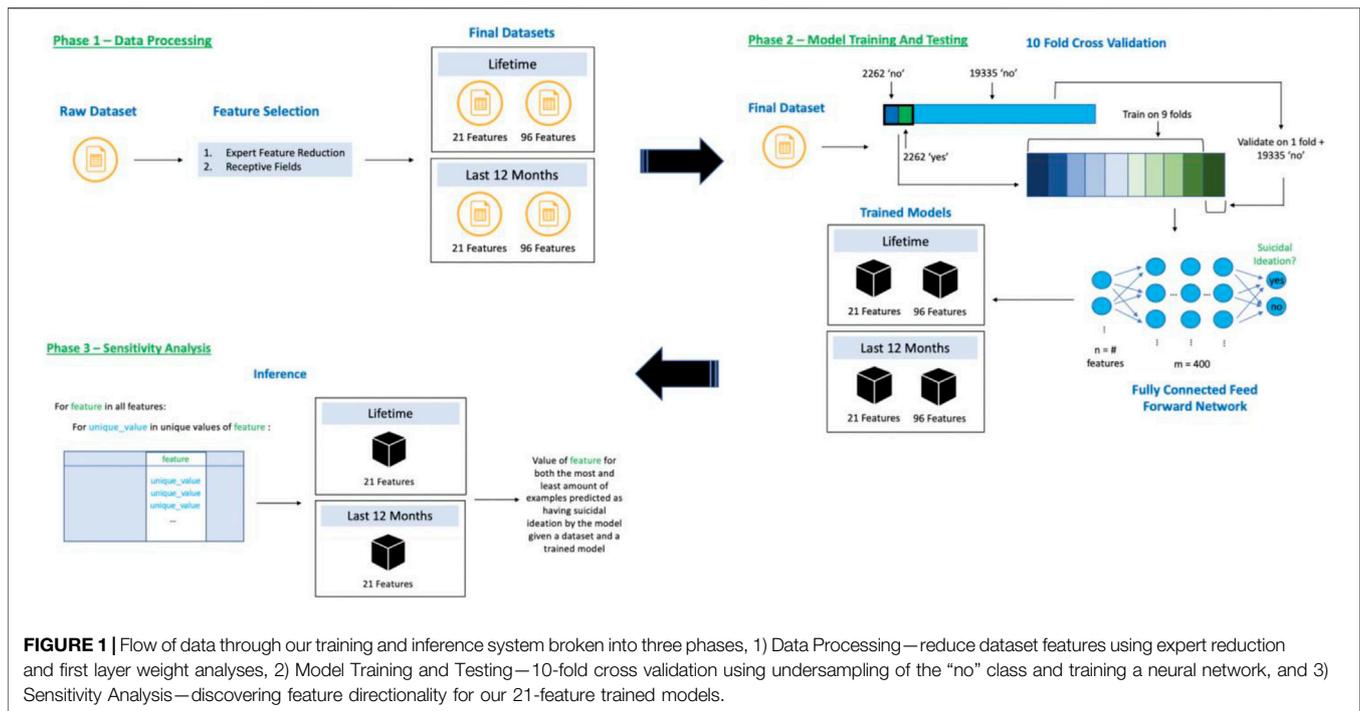
In order to obtain a model that could be implemented in a real clinical environment, reducing the number of input features to

pinpoint the most important features in the dataset is necessary. A model requiring too many input features would present challenges for data collection in the clinic when, making it difficult to apply the model to a given patient rapidly and efficiently¹. The techniques used for feature selection involved both expertise in the field (i.e., expert feature reduction) and allowing the model to highlight which features were the most important (Guyon and Elisseeff, 2003). A clinician (D. B.) went through all 582 features and discarded the features which were either administrative (i.e., redundant case identification codes or different ways of asking the same question) or which were not reasonable to collect clinically (such as detailed health care service satisfaction metrics which would not be appropriate in a screening context where the patient has not yet experienced services fully). This reduced the feature set size to 196. We further reduced the number of features using machine learning techniques. This involved analyzing the weights fields of the trained model’s first layer and removing “unimportant” features. Feature “importance” was defined *via* the weights that the neural network applied to a particular feature, which was inspired by the concept of receptive fields in convolutional networks (Coates and Ng, 2011). Our motivation for looking at the first layer weights is that if the model effectively “drops out” that feature by masking it with a near-zero value, it will not play a role in further nonlinear interactions. This assumption can only be made for the first layer since any following intermediate layer contains too many complex interactions, precluding the association of a direct weight to a feature. We cannot make the converse assumption that high values at the first layer will equate to high feature importance because of nonlinear interactions which occur at intermediate layers. A visualization of this feature selection technique is shown in **Supplementary Figure S2**.

Two cases were examined, one in which 100 features were removed, leaving 96 features in the model, and one in which 175 features were removed, leaving 21 features in the model. We chose to remove 100 and 175 features respectively, since the 100 feature removal didn’t affect the performance too much from the larger feature set sizes (>100 features) and stopped at 175 because removing any more features would cause the performance to deteriorate. The larger models were produced in order to maximize the identification of important features and to maximize model accuracy; the smaller models were produced in order to generate clinically tractable models with few enough questions that they could be integrated into a standard screening assessment. Separate models were produced for both lifetime and last-12-months suicidal ideation identification.

In order to adjust our model to the large class imbalance that existed between the “no” and “yes” responders, we used undersampling. The number of examples in the majority (“no”) class was equated to the number in the minority (“yes”) class. In the case of lifetime identification, 2,262 random examples from the “no” class were randomly chosen

¹A note on terminology: “feature” here refers to an input variable (i.e., one survey item).



for the training set to match the 2,262 samples from the “yes” class. The class-balanced training set was then divided into 10 different random folds, and the model was trained on nine of these folds, leaving the final fold and all of the other 19,355 “no”s to serve as the validation set. This process was repeated 10 times with mutually exclusive validation and training sets, and we noted the average of the test metrics of all runs on the validation set. It is important to mention here that our validation set was comprised of a relatively lower count of respondents in the “yes” class compared to the initial distribution of the data, making it much harder for the model to be able to classify respondents in the “yes” class correctly. The same sort of division was performed for the last-12-months data using the data distribution shown in **Supplementary Table S6**.

All analyses were done using the Vulcan software package (see software note). **Figure 1** represents the steps taken to produce the results for this analysis.

RESULTS

Supplementary Appendix Tables A1–A4 show the features used for the identification of lifetime suicidal ideation (**Supplementary Appendix Table A1**: 96 features, **Supplementary Appendix Table A2**: 21 features) and suicidal ideation during the past 12 months (**Supplementary Appendix Table A3**: 96 features, **Supplementary Appendix Table A4**: 21 features). These features are those that remain following expert feature reduction (manual feature removal using domain expertise) and using the network’s first layer weights to remove additional features until 96 and 21 features remained for both lifetime and last-12-months suicidal ideation models. In terms of measurement, we chose to use the

Area under the receiver operating curve (AUC) as our main metric of model performance, and we also calculated the sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) for each model. **Supplementary Appendix Table A5** shows the 10-fold cross validated results for the lifetime (96 features—0.7890 AUC; 21 features—0.7681 AUC) and last-12-months (96 features—0.7081 AUC; 21 features—0.6798 AUC) datasets, respectively. Random forest and gradient boosted tree classifiers were produced as non-deep learning baselines; these generally performed quite well (**Supplementary Appendix Table A5**), with gradient boosted trees having the best metrics across all four models and random forest doing worse than deep learning in the lifetime models, but better in the last-12-months models; deep learning suffers as the training set size decreases and the last-12-months dataset is much smaller than the lifetime dataset (**Supplementary Table S6**). Note that the purpose of this paper was not to demonstrate superiority of deep learning, but rather to establish its feasibility and potential interpretability in this type of dataset, with the intention of using deep learning in larger datasets where, as we discuss below, it is likely to outperform other model types. In total, we produced four model configurations: 96 and 21 features for identification of lifetime suicidal ideation and 96 and 21 features for identification of suicidal ideation in the last 12 months.

In order to gain insight into how different features affected model classifications (i.e., feature directionality) and improve interpretability, we performed a feature sensitivity analysis for the 21-feature models. We chose not to perform the same analysis for the 96-feature models as it would be unsuitable to interpret due to size. We explored how variations in values for a specific feature affected the final model classification. We accomplished this by iterating through all possible unique values (up to a

maximum of 20 values) for each feature and imputed all response samples to have this value. We then ran a test to determine how many of the samples would be classified as having suicidal ideation by the model. The third columns in the 21 feature tables (**Supplementary Appendix Tables A2, A4**) show the value of the feature where the model predicted the highest amount of suicidal ideation followed by the feature value with the lowest amount of suicidal ideation. In **Supplementary Appendix Tables A2, A4**, the number in brackets next to each feature value shows the number of examples in the test set classified as having suicidal ideation (19,788 samples in the test set for lifetime; 1,769 for the past 12 months). This allows for some insight into the inner workings of the neural network model. For example, in the lifetime identification of suicidal ideation, if all the answers to the question “have people to count on in an emergency” are set to “strongly disagree”, then 8,046 people are identified as having suicidal ideation; this number drops to 5,158 people if the answers are all changed to “strongly agree”.

In order to investigate model performance in subgroups which may be relevant to the prediction of suicidal ideation, we ran each of the four models three times on different splits of data, each time examining the AUCs achieved in each of a number of subgroups (these subgroups were defined by gender, household income, and education). The results can be seen in **Supplementary Table S9**.

DISCUSSION

This initial proof of concept illustrates that using our method, data from the general population can help identify people at risk of suicidal ideation. These people might benefit from more in-depth screening and resources in the context of suicide prevention, and further work using these kinds of methods might contribute to the development of clinically useful screening tools.

Jordan et al. (2018) found that using only four items of the PHQ-9 provided the most accurate predictions of suicidal ideation in their patient sample—those assessing “feelings of depression/hopelessness, low self-esteem, worrying, and severe sleep disturbances” (Jordan et al., 2018). Although the PHQ-9 was not included in our dataset, our model similarly found some high impact variables related to depression and worrying. For instance, having generalized anxiety disorder or depression were associated with suicidal ideation in our model (**Supplementary Appendix Tables A1, A2**). Unlike the Jordan model, ours did not identify sleep problems to be a significant risk factor for suicidal ideation. One possible explanation reconciling our results and those in the literature is that sleep problems may act as a proxy for actual interacting risk factors rather than being a risk factor themselves. When such factors are included in the data and processed by a complex model, sleep disorder factors are rendered irrelevant. We will seek to verify this hypothesis in other datasets with more robust measures of sleep. While Jordan’s model identified low self-esteem as a risk factor, our dataset unfortunately did not contain a self-esteem variable. Our model

yielded additional predictive factors that do not overlap with those found by the Jordan team. Generalized anxiety disorder, for example, appears to be an important predictor of suicidal ideation (**Supplementary Appendix Tables A1–A3**). This is to be expected, since previous research has identified anxiety disorders, including generalized anxiety disorder, as independently predictive of suicidal ideation (Sareen et al., 2005; Bentley et al., 2016). Importantly, our method yielded predictors related to early traumatic experiences and diagnosis of PTSD. Non-consensual sexual experiences before the age of 16, appear to be associated with suicidal ideation, as is early physical abuse and number of types of childhood trauma experienced (**Supplementary Appendix Tables A1–A4**). This finding is supported by previous research linking increased suicidal ideation and suicide attempts to early sexual abuse, and early physical abuse to suicidal ideation through an association with anxiety, which was also an important factor identified by the model (Basile et al., 2006; Ullman et al., 2009; Bedi et al., 2011; Lopez-Castroman et al., 2013; Bahk et al., 2017; Thompson et al., 2018), thus confirming our model’s capacity to identify known risk factors of suicidal ideation. There is extensive literature suggesting that early-life adversity is an important predictor of suicidal behavior (Brezo et al., 2008; Wanner et al., 2012; Turecki and Brent, 2016). In fact, adverse childhood experiences were demonstrated to account for 67% of the population attributable risk for suicide attempts (Teicher et al., 2016).

It is interesting to note that some factors, such as generalized anxiety disorder in the lifetime 21-features model, do not have a clear directionality (i.e., they have a ratio in the sensitivity analysis close to one), and yet are included in the model as important risk or protective factors. This may be because these factors interact with other factors to produce their effect. For example in the same model physical abuse in childhood was strongly associated with SI. This opens the possibility that we are replicating the results reported by Bahk et al. (2017), where childhood physical abuse was predictive of suicidal ideation through an association with anxiety.

An important aspect of our model is potential for generalizability to a real population. We note that another paper, by Ryu et al. (2018), built a random forest classifier for the identification of suicidal ideation in a general population sample. Their approach was to downsample their data to achieve a 1:1 ratio between those endorsing and not endorsing suicidal ideation, even in their test set. Our work improves on this approach as our validation fold included a significant, more realistic, class imbalance with those not reporting suicidal ideation vastly outnumbering those who did and our results being representative of how such a model would perform when applied to a population where a class imbalance exists. As such, our model may be more generalizable to a real population.

Like ours, the Ryu model (2018) found depression and anxiety to be some of the most important features predicting SI. Sociodemographic features such as age, sex, education and features related to quality of life and employment were also predictive of SI in both models. Medical comorbidities were predictive in both models, however Ryu also noted that somatic symptoms predicted SI.

We separated identification of suicidal ideation occurring in the last 12 months and throughout the lifetime to disambiguate more specific short-term from long-term predictors. Identification of protective factors and risk factors for both conditions may improve methods of identifying and treating those at risk of attempting suicide. Lifetime factors may be useful in developing more long-term suicide prevention strategies, while factors identifying suicide ideation in the last 12 months can inform the identification and treatment of patients at more immediate risk. While all predictors were related to physical and mental wellbeing, mental health, early abusive experiences, socioeconomic situation and social support, some differences between risk factors and protective factors for the lifetime and last 12 months conditions may be important to consider. For example, having a trustworthy person to turn to for advice, seems to have significant weight in the last 12 months model. This may indicate that measures of social support, specifically of close support relationships, could be used to identify patients at more immediate risk of suicidal ideation. Based on previous literature, lack of social support may be a moderator between life stress and suicidal ideation, suggesting that a strong social support system may be beneficial in reducing suicidal thoughts, particularly during stressful times (Yang and Clum, 1994; Vanderhorst and McLaren, 2005); this seems to be relevant to both the last-12-months and lifetime models.

Additionally, contact with the police was present in the expanded models and may be associated with suicidal ideation, highlighting a need to follow up with people who may have had a traumatic experience leading to police intervention, or negative interactions with the police (DeVylder et al., 2018).

We identified several predictors that are easy to obtain, including sociodemographic features. Interestingly, Jordan et al. did not find sociodemographic features useful in the prediction of suicidal ideation (2018), but as we were using a census dataset with a large and varied array of sociodemographic features, we were able to identify more predictors amongst them that would be amenable to upstream intervention. As opposed to more expensive data like neuroimaging and genetic testing, sociodemographic predictors can be very useful in clinical practice, especially with respect to screening, since they are easily accessible to healthcare professionals through direct questioning or self-report questionnaires.

As can be seen in **Supplementary Appendix Table A5**, the 96-feature models have higher AUCs. This is to be expected, as the network is able to make better identifications when it has more information on the different patients it is ingesting. It is worthwhile to discuss the pros and cons of having larger or smaller models. Large models that do not overfit allow us to identify more predictors, which may be modifiable and are therefore potentially useful from a public health standpoint. Smaller models are easier to implement because patients need to answer fewer questions in order to provide the model with sufficient information to make a prediction. Thus, there exists an interesting trade-off between model accuracy and ease of data acquisition when selecting the number of features to include. For example, the difference in the AUC for the last-12-months model presented here is 0.68 for the 21-feature model vs. 0.71 for the 96-feature model. Does this difference justify a larger model that is

more accurate but more difficult to collect? while the difference between the two AUC values may seem insignificant, when considering predictions on a population scale we might expect a significant difference in the absolute number of people correctly classified. Implementation of models such as these will hinge on finding the right balance between model complexity and accuracy in order to provide models that are both meaningful and feasible to implement.

It is also important to note the high negative predictive values (NPV) of our predictions. This metric indicates that the network is almost always correct when it classifies an example as not having suicidal ideation. This is crucial, as it signifies one potential use of our model may be in helping, alongside good clinical judgement and history taking, to rule out suicidal ideation in populations matching those in the dataset. However, it should be noted that given the rarity of suicidal ideations in this population, a high NPV is to be expected given that “not suicidal” is the dominant prediction class (Belsher et al., 2019)]. Given that clinicians currently have difficulty ruling out suicidal thinking or risk (McDowell et al., 2011), such a tool could eventually be clinically helpful if the rate of false negatives is judged to be acceptable. Further research is required to determine an acceptable false negative rate and how to better integrate these kinds of models into clinical decision making. Concerns about false negatives must be balanced against the risk of false positives, which can lead to unnecessary intervention and confinement, as well as against the fact that the absence of suicidal ideation at a single point in time does not rule out the risk of suicide (McHugh et al., 2019). However, given that this model identifies suicidal ideation and not risk of attempt, a positive result could be used to open a conversation between a clinician and patient, which might lead to more appropriate assessment and treatment before the risk of an attempt increases. This in turn may become a useful approach for the prevention of suicide *via* upstream identification of at-risk patients in the general population, though this remains speculative and should be expanded on in future work exploring factors that predict conversion of ideation to action. Given the PPV values we report [which are low but in line with the literature (Belsher et al., 2019)], it would be crucial that any interventions developed using similarly performing models be carefully designed in order to favor low-intensity interventions such as further assessment by a clinician over any restrictive or high-intensity interventions. Alternatively, future work could focus on improving the PPV and other metrics of these models.

There are several limitations to our current work. While using an interview-based census dataset allows for a large sample size in the general population, it does mean that there are no clinician-rated scales or independent verification of participant responses. Our use of deep learning, intended here as a proof of concept, provides a powerful technique that has the potential to match or outperform other methods as the dataset size increases (Alom et al., 2019). While deep learning is generally less easy to interpret than other machine learning techniques, our sensitivity analysis does allow some insight into the model parameters which could be further evaluated using classical statistics. This work is an initial proof of concept that could lead to a more comprehensive solution

that would identify suicidal ideation and suicidal behaviours with increased reliability. While gradient boosting currently outperforms deep learning in this dataset, deep learning performed well and the sensitivity analysis renders it more interpretable, which in turn makes further study of deep learning models in larger datasets attractive. In addition, the sensitivity analysis may be useful in determining when a model which may have reasonably good metrics may have learned inappropriate relationships between features, potentially limiting generalizability to other datasets or to specific subsets of patients. Examining the results of a sensitivity analysis may uncover potentially spurious relationships learned by a model when the underlying data has an unexpected skew. Thus, when working with imbalanced datasets that require different sampling approaches [like the one in this paper], or when working with deep learning in datasets which are expected to contain known relationships between features, the sensitivity analysis may be a useful tool beyond model metrics in examining model quality and likely generalizability. A significant limitation to this current paper is the lack of an independent dataset for validation, notwithstanding our cross-validation approach in which test splits were designed to pose a harder problem than training splits by nature of the greater class imbalance in the test splits. Future work will need to test this model, or future successor models, on independent datasets to ensure generalizability.

We note that the data used in this paper had a significant breadth for each subject, with many different symptoms and socio-demographic and economic factors explored for each person included. However this comes at a cost of depth and specificity, with many features not necessarily containing all the information one might hope for. For example, our target feature—suicidal ideation—is a simple binary variable that does not include information about intensity, frequency or recurrence. In addition, it is not possible to verify the accuracy of each label of “yes” or “no” for suicidal ideation, with many “no”s potentially being from those who had ideations in the past and had forgotten, or who wished not to reveal ideations to a census agent. Despite this, the features identified by the model do seem reasonable when considered in the light of previous suicidal ideation literature, which speaks to the fact that the model seems to be picking up on realistic predictors. In addition, it must be remembered that this model is primarily a proof of concept, not a clinic-ready model, and will require further validation on different kinds of datasets prior to being ready for clinical testing.

One significant limitation of the model is that it identifies suicidal ideations rather than suicide attempt or completion risk, and as such its clinical utility is limited. For patients already being assessed by a mental health professional, or being interviewed for the purpose of determining SI, it is far simpler to simply ask them about lifetime or recent SI than to rely on or collect the data required to generate a model result. However, given that the data used to train this model was derived from the general population, it may have utility in the screening of patients in general practice, or *via* automatic review of electronic medical records. This could

prompt clinicians to engage patients who they may not otherwise think to ask about SI, leading to more patients being referred to appropriate services as needed. In addition, the features identified in the larger 96 feature models may be helpful to those setting healthcare and social services priorities when considering interventions aimed at addressing or decreasing suicidal ideation. Prior to the deployment of any such model in the clinical setting further clinical research would be required.

Another relevant test for future models, trained on more examples of persons with suicidal ideation, would be to see if the model accurately recapitulates risk of suicidal ideation in specific strata of a risk-stratified population. In an effort to investigate how this set of models would perform in subgroups, we ran them three times each on different splits of data in order to see if model accuracy suffered when looking at subgroups within potentially relevant subject categories (for example, if the model had worse performance when used in only female subjects, or only people with a certain socioeconomic status). These results, detailed in **Supplementary Table S9**, demonstrate that the model performs well in all subgroups for which there was a reasonable amount of data. Future work should see models trained with and tested on larger proportions of patients from relevant subgroups in order to ensure that groups at risk of suicidal ideation, or traditionally underserved groups, are well represented in the model.

It is worth discussing the practical implementation of a tool for the identification of suicidal ideation in clinical practice, as this would bring both possible benefits and challenges. One possible implementation of this tool would be as an automated screening tool integrated into electronic medical records in emergency departments or outpatient clinics. Benefits—which would need to be verified in clinical studies—could include earlier and more accurate identification of suicidal ideation, which would lead to more patients being offered appropriate services, such as access to a therapist or to crisis resources. This in turn would hopefully lead to a reduction in the number of patients making suicide attempts or completing suicide, though this would depend on the efficacy of the offered interventions. Nonetheless, challenges and potential dangers exist. Models that identify suicidal ideation could be used by some clinicians to justify interventions such as forced hospitalization, which raises serious concerns about the effect of implementing such models on patient autonomy and clinician medico-legal risk. In addition, it is unclear what effect having an automated screening tool for suicidal ideation would have on clinician behavior. It might improve clinician awareness of the importance of screening for and offering support to patients with suicidal ideation. At the same time, since the model cannot currently identify all patients with SI or rule out all patients without SI, it may reinforce the habit of many clinicians to avoid asking about suicidal ideation, fostering an over-reliance on an imperfect system to screen for a potentially serious clinical phenomenon. Any implementation of such a screening system would require significant investment in the training of clinicians and should be accomplished in partnership with patient and clinician representatives.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www150.statcan.gc.ca/n1/en/catalogue/82M0013X> <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&id=119789>. Software Note: the Vulcan Platform can be accessed at <https://github.com/AifredHealth/Vulcan>.

AUTHOR CONTRIBUTIONS

SD carried out the experiment and helped prepare the manuscript. MT-S helped prepare and revise the manuscript and conducted a literature review. DB helped conceptualize the experiment, prepare and revise the manuscript. RF helped

conceptualize and review the experiment, provided the computational tools, and helped revise the manuscript. EB and KP helped conduct a literature review and revise the manuscript. AJ and MD-B assisted in the revision of the manuscript and in the revision of the experiment planning and results presentation. NL, SI, and LP assisted in revising the manuscript and completing the literature review. GT revised the manuscript and provided supervision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2021.561528/full#supplementary-material>

REFERENCES

- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., et al. (2019). A State-Of-The-Art Survey on Deep Learning Theory and Architectures. *Electronics*, 8, 292. doi:10.3390/electronics8030292
- Bahk, Y.-C., Jang, S.-K., Choi, K.-H., and Lee, S.-H. (2017). The Relationship between Childhood Trauma and Suicidal Ideation: Role of Maltreatment and Potential Mediators. *Psychiatry Investig.* 14 (1), 37–43. doi:10.4306/pi.2017.14.1.37
- Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., et al. (2016). Predicting Suicidal Behavior from Longitudinal Electronic Health Records. *Am. J. Psychiatry*, 174 (2), 154–162. doi:10.1176/appi.ajp.2016.16010077
- Basile, K. C., Black, M. C., Simon, T. R., Arias, I., Brener, N. D., and Saltzman, L. E. (2006). The Association between Self-Reported Lifetime History of Forced Sexual Intercourse and Recent Health-Risk Behaviors: Findings from the 2003 National Youth Risk Behavior Survey. *J. Adolesc. Health*, 39 (5), e1–752. doi:10.1016/j.jadohealth.2006.06.001
- Beck, A. T., Kovacs, M., and Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *J. Consult. Clin. Psychol.* 47 (2), 343–352. doi:10.1037/0022-006x.47.2.343
- Bedi, S., Nelson, E. C., Lynskey, M. T., Mc Cutcheon, V. V., Heath, A. C., Madden, P. A. F., et al. (2011). Risk for Suicidal Thoughts and Behavior after Childhood Sexual Abuse in Women and Men. *Suicide Life-Threatening Behav.* 41 (4), 406–415. doi:10.1111/j.1943-278x.2011.00040.x
- Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., et al. (2019). Prediction Models for Suicide Attempts and Deaths. *JAMA Psychiatry*, 76 (6), 642–651. doi:10.1001/jamapsychiatry.2019.0174
- Bentley, K. H., Franklin, J. C., Ribeiro, J. D., Kleiman, E. M., Fox, K. R., and Nock, M. K. (2016). Anxiety and its Disorders as Risk Factors for Suicidal Thoughts and Behaviors: A Meta-Analytic Review. *Clin. Psychol. Rev.* 43, 30–46. doi:10.1016/j.cpr.2015.11.008
- Bongiovi-Garcia, M. E., Merville, J., Almeida, M. G., Burke, A., Ellis, S., Stanley, B. H., et al. (2009). Comparison of Clinical and Research Assessments of Diagnosis, Suicide Attempt History and Suicidal Ideation in Major Depression. *J. Affective Disord.* 115 (0), 183–188. doi:10.1016/j.jad.2008.07.026
- Breiman, L. (2001). Random Forests. *Machine Learn.* 45 (1), 5–32. doi:10.1023/a:101793452217110.1023/a:1010933404324
- Brezo, J., Paris, J., Vitaro, F., Hébert, M., Tremblay, R. E., and Turecki, G. (2008). Predicting Suicide Attempts in Young Adults with Histories of Childhood Abuse. *Br. J. Psychiatry*, 193 (2), 134–139. doi:10.1192/bjp.bp.107.037994
- Cai, L., Wang, Z., Gao, H., Shen, D., and Ji, S. (2018). Deep Adversarial Learning for Multi-Modality Missing Data Completion. *Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discov. Data Mining - KDD*, 18, 1158–1166. doi:10.1145/3219819.3219963
- Coates, A., and Ng, A. Y. (2011). Selecting Receptive Fields in Deep Networks. *NIPS*, 2528–2536. doi:10.1016/j.psychres.2009.03.008
- Delpozo-Banos, M., John, A., Petkov, N., Berridge, D. M., Southern, K., Lloyd, K., et al. (2018). Using Neural Networks with Routine Health Records to Identify Suicide Risk: Feasibility Study (Preprint). *JMIR Ment. Health*, 5 (2). doi:10.2196/preprints.10144
- DeVylder, J. E., Jun, H.-J., Fedina, L., Coleman, D., Anglin, D., Cogburn, C., et al. (2018). Association of Exposure to Police Violence with Prevalence of Mental Health Symptoms Among Urban Residents in the United States. *JAMA Netw. Open* 1 (7), e184945. doi:10.1001/jamanetworkopen.2018.4945
- Exarchos, K. P., Goletsis, Y., and Fotiadis, D. I. (2012). Multiparametric Decision Support System for the Prediction of Oral Cancer Recurrence. *IEEE Trans. Inform. Technol. Biomed.* 16 (6), 1127–1134. doi:10.1109/TITB.2011.2165076
- Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., et al. (2017). Risk Factors for Suicidal Thoughts and Behaviors: A Meta-Analysis of 50 Years of Research. *Psychol. Bull.* 143 (2), 187–232. doi:10.1037/bul0000084
- Guyon, I., and Elisseeff, A. (2003). “An Introduction to Variable and Feature Selection,” Editor L. P. Kaelbling *J. Machine Learn. Res.*, 3, 1157–1182.
- Hubers, A. A. M., Moaddine, S., Peersmann, S. H. M., Stijnen, T., van Duijn, E., van der Mast, R. C., et al. (2016). Suicidal Ideation and Subsequent Completed Suicide in Both Psychiatric and Non-psychiatric Populations: A Meta-Analysis. *Epidemiol. Psychiatr. Sci.* 27 (2), 186–198. doi:10.1017/s20457960160001049
- Jordan, P., Shedden-Mora, M. C., and Löwe, B. (2018). Predicting Suicidal Ideation in Primary Care: An Approach to Identify Easily Assessable Key Variables. *Gen. Hosp. Psychiatry*, 51, 106–111. doi:10.1016/j.genhosppsych.2018.02.002
- Kingma, D. P., and Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. In International Conference For Learning Representations.
- Klambauer, G., Unterthiner, T., and Mayr, A. (2017). “Self-normalizing Neural Networks,” in *Advances in Neural Information Processing Systems*. I. Guyon, U. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. Editors (Red Hook, NY: NIPS), 30.
- Kroenke, K., and Spitzer, R. L. (2002). The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr. Ann.* 32 (9), 509–515. doi:10.3928/0048-5713-20020901-06
- Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. General Int. Med.* 16 (9), 606–613. doi:10.1046/j.1525-1497.2001.016009606.x
- Lopez-Castroman, J., Melhem, N., Birmaher, B., Greenhill, L., Kolko, D., Stanley, B., et al. (2013). Early Childhood Sexual Abuse Increases Suicidal Intent. *World Psychiatry*, 12 (2), 149–154. doi:10.1002/wps.20039
- Manaktala, S., and Claypool, S. R. (2017). Evaluating the Impact of a Computerized Surveillance Algorithm and Decision Support System on Sepsis Mortality. *J. Am. Med. Inform. Assoc.* 24 (1), 88–95. doi:10.1093/jamia/ocw056

- McDowell, A. K., Lineberry, T. W., and Bostwick, J. M. (2011). Practical Suicide-Risk Management for the Busy Primary Care Physician. *Mayo Clinic Proc.* 86 (8), 792–800. doi:10.4065/mcp.2011.0076
- McHugh, C. M., Corderoy, A., Ryan, C. J., Hickie, I. B., and Large, M. M. (2019). Association between Suicidal Ideation and Suicide: Meta-Analyses of Odds Ratios, Sensitivity, Specificity and Positive Predictive Value. *BJPsych Open.* 5 (2). doi:10.1192/bjo.2018.88
- Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M.-J., Zhang, X. Y., et al. (2016). Identifying a Clinical Signature of Suicidality Among Patients with Mood Disorders: A Pilot Study Using a Machine Learning Approach. *J. Affective Disord.* 193, 109–116. doi:10.1016/j.jad.2015.12.066
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Machine Learn. Res.* 12, 2825–2830.
- Ryu, S., Lee, H., Lee, D.-K., and Park, K. (2018). Use of a Machine Learning Algorithm to Predict Individuals with Suicide Ideation in the General Population. *Psychiatry Investig.* 15 (11), 1030–1036. doi:10.30773/pi.2018.08.27
- Sareen, J., Cox, B. J., Afifi, T. O., de Graaf, R., Asmundson, G. J. G., ten Have, M., et al. (2005). Anxiety Disorders and Risk for Suicidal Ideation and Suicide Attempts. *Arch. Gen. Psychiatry.* 62 (11), 1249–1257. doi:10.1001/archpsyc.62.11.1249
- Srivastava, A. S., and Kumar, R. (2005). Suicidal ideation and attempts in patients with major depression: Sociodemographic and clinical variables. *Indian J. Psychiatry.* 47 (4), 225. doi:10.4103/0019-5545.43059
- Teicher, M. H., Samson, J. A., Anderson, C. M., and Ohashi, K. (2016). The Effects of Childhood Maltreatment on Brain Structure, Function and Connectivity. *Nat. Rev. Neurosci.* 17, 652–666. doi:10.1038/nrn.2016.111
- Thompson, M. P., Kingree, J. B., and Lamis, D. (2018). Associations of Adverse Childhood Experiences and Suicidal Behaviors in Adulthood in a U.S. Nationally Representative Sample. *Child. Care Health Dev.* 45 (1), 121–128. doi:10.1111/cch.12617
- Turecki, G., and Brent, D. A. (2016). Suicide and Suicidal Behaviour. *The Lancet.* 387 (10024), 1227–1239. doi:10.1016/S0140-6736(15)00234-2
- Ullman, S. E., and Najdowski, C. J. (2009). Correlates of Serious Suicidal Ideation and Attempts in Female Adult Sexual Assault Survivors. *Suicide Life-Threatening Behav.* 39 (1), 47–57. doi:10.1521/suli.2009.39.1.47
- Vanderhorst, R. K., and McLaren, S. (2005). Social Relationships as Predictors of Depression and Suicidal Ideation in Older Adults. *Aging Ment. Health.* 9 (6), 517–525. doi:10.1080/13607860500193062
- Walsh, C. G., Ribeiro, J. D., and Franklin, J. C. (2017). Predicting Risk of Suicide Attempts over Time through Machine Learning. *Clin. Psychol. Sci.* 5 (3), 457–469. doi:10.1177/2167702617691560
- Wanner, B., Vitaro, F., Tremblay, R. E., and Turecki, G. (2012). Childhood Trajectories of Anxiousness and Disruptiveness Explain the Association between Early-Life Adversity and Attempted Suicide. *Psychol. Med.* 42 (11), 2373–2382. doi:10.1017/s0033291712000438
- Yang, B., and Clum, G. A. (1994). Life Stress, Social Support, and Problem-Solving Skills Predictive of Depressive Symptoms, Hopelessness, and Suicide Ideation in an Asian Student Population: A Test of a Model. *Suicide Life Threat Behav.* 24, 127–139. doi:10.1111/j.1943-278X.1994.tb00797.x

Conflict of Interest: MT-S and SI are employees and shareholders of Aifred Health, a medical technology company that uses deep learning to increase treatment efficacy in psychiatry. DB, RF and KP are shareholders and/or employees of Aifred Health. SD and EB have received consulting fees from Aifred Health.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Desai, Tanguay-Sela, Benrimoh, Fratila, Brown, Perlman, John, DelPozo-Banos, Low, Israel, Palladini and Turecki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.