Northeast Journal of Complex Systems (NEJCS)

Volume 4
Number 1 *Special Issue: NERCCS 2022 Papers*

Article 2

March 2022

Toward Suicidal Ideation Detection with Lexical Network Features and Machine Learning

Ulya Bayram
Çanakkale Onsekiz Mart University, ulya.bayram@comu.edu.tr

William Lee
University of Massachusetts Amherst, williamlee@umass.edu

Daniel Santel
Cincinnati Children's Hospital Medical Center, Daniel.Santel@cchmc.org

Ali Minai
University of Cincinnati, ali.minai@uc.edu

Peggy Clark
Cincinnati Children's Hospital Medical Center, Peggy.Clark@cchmc.org

Toward Suicidal Ideation Detection with Lexical Network Features and Machine Learning

Authors

Ulya Bayram, William Lee, Daniel Santel, Ali Minai, Peggy Clark, Tracy Glauser, and John Pestian

Toward Suicidal Ideation Detection with Lexical Network Features and Machine Learning

Ulya Bayram, Ph.D.[1*], William Lee[2], Daniel Santel, Ph.D.[3,4], Ali A. Minai, Ph.D.[5], Peggy O. Clark, DNP, MSN, PPCNP-BC[4], Tracy Glauser, MD[3,4], and John Pestian, Ph.D.[3,6]

[1] Department of Electrical and Electronics Engineering, Faculty of Engineering, Çanakkale Onsekiz Mart University, Çanakkale, Turkey
[2] Department of Computer Science, University of Massachusetts Amherst, Massachusetts, USA
[3] Department of Pediatrics, Divisions of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati Ohio, USA
[4] Department of Pediatrics, Divisions of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati Ohio, USA
[5] Department of Electrical Engineering and Computer Science, Faculty of Engineering, University of Cincinnati, Cincinnati Ohio, USA
[6] Department of Pediatrics, Divisions of Psychiatry, Cincinnati Children's Hospital Medical Center, Cincinnati Ohio, USA
[*] ulya.bayram@comu.edu.tr

Abstract

In this study, we introduce a new network feature for detecting suicidal ideation from clinical texts and conduct various additional experiments to enrich the state of knowledge. We evaluate statistical features with and without stopwords, use lexical networks for feature extraction and classification, and compare the results with standard machine learning methods using a logistic classifier, a neural network, and a deep learning method. We utilize three text collections. The first two contain transcriptions of interviews conducted by experts with suicidal (n=161 patients that experienced severe ideation) and control subjects (n=153). The third collection consists of interviews conducted by experts with epilepsy patients, with a few of them admitting to experiencing suicidal ideation in the past (32 suicidal and 77 control). The selected methods detect suicidal ideation with an average area under the curve (AUC) score of 95% on the merged collection with high suicidal ideation, and the trained models generalize over the third collection with an average AUC score of

69%. Results reveal that lexical networks are promising for classification and feature extraction as successful as the deep learning model. We also observe that a logistic classifier's performance was comparable with the deep learning method while promising explainability.

1 Introduction

Suicide and its prevention is a problem of growing importance according to the increasing global suicide statistics [1, 2]. Meanwhile, the aftermath of the COVID-19 pandemic on suicide rates has started to surface the dire effects of social isolation, lock-downs, stress, and anxiety factors affecting mental health and triggering suicidal events [3, 4, 5]. Approximately 70% of the suicidal individuals needing urgent mental health services remain helpless because of the shortage of caregivers, lack of health insurance, or by choice [6, 7]. Meanwhile, accessing these services cannot guarantee suicide prevention either. Healthcare providers without good training who solely rely on standardized questionnaires may fail to see the nuances of suicide [8, 9, 10]. Also, suicidal individuals often deny and hide their suicidal thoughts [11]. A solution considered by many has been to adopt machine learning technology for suicide prevention [12].

Many studies evaluate the possibility of identifying suicidal ideation from texts using machine learning [13, 14, 15, 16]. Deep learning methods such as convolutional neural networks (CNN) [17, 18], long-short term memory networks (LSTM) [17, 19], and BERT models [20] have also been used for identifying the presence of suicidal ideation. The reported scores reach as high as above 90% AUC over within-corpus evaluations [8, 21]. Due to the reported success of these methods, some social media domains already started to use such models in real-time detection and prevention of suicide [12]. These studies made it possible to demonstrate the plausibility of suicidal ideation detection using supervised models.

Despite the developments in the machine learning applications, the current state of research is far from completion. There are only a few studies that attempt to utilize complex networks in the task of text classification [22, 23], especially in the mental health domain. Also, there is an insufficiency of studies that use expert-labeled data collections due to the expensive and difficult nature of collecting clinical data. Plus, there is a need for studies that perform fair method comparisons instead of reporting results from a single approach or perform simple comparisons with methods that were not properly tuned. Most importantly, literature needs studies that report how the trained models with high within-dataset results would generalize over other collections and studies that demonstrate the interpretability of the methods. In this study, we respond to these demands.

In this study, we use three clinical data collections available to us. The first two

collections comprise interview transcripts of individuals in the hospital due to suicidal events and control individuals. Meantime, the third collection has interview transcripts of patients that received epilepsy treatment in the past, some of them admitting having suicidal ideation currently or in the past. We use the third one for studying the generalization/transferability of the machine learning methods trained by a subset of the first two collections, focusing on determining the generalization of the high-scoring models in a real-life scenario. The significance of this third collection is the low-level suicidal ideation and the small number of suicidal individuals it contains compared to the first two, making it a realistic experimentation framework.

In the following sections, we discuss our contributions, summarize the methods, demonstrate and discuss the results, and finalize the main findings in the conclusions.

2 Related Work

The majority of the suicidal ideation detection research domain is dominated by machine learning and deep learning studies [13, 14, 15, 16, 8]. Using networks for analyzing or identifying suicidal ideation have been covered by a few studies. In one study, De et al. constructs a network to analyze the interactions between different theoretical components of suicidal ideation [24]. In addition to the theoretical studies, there are others that focus on utilizing the power of cognitive network science. One recent study constructs a lexical network from the suicide notes to study and analyze the associations between the terms in these notes [23]. They find clusters of positive words dominating their networks, while they find negative terms highly clustered around the self-related terms in the networks. As such analyses provide a grasp of the strength of constructing networks from the data, some studies focused on using this strength to distinguish a suicidal texts from the others, using the networks as a classifier [22]. In this study, we introduce lexical networks as a resourceful domain for feature extraction. The rationale is that these features reflect more subtle lexical relationships than are captured by standard statistical features (e.g., bi-grams) and can be used to augment or replace these features in machine learning applications for detecting suicidal ideation. Additionally, we introduce an analysis on the effects of handling stopwords, i.e. frequent words that are meaningless alone, in the feature set as there is no consensus regarding their removal. Next, we perform a feature analysis to address the issue regarding the lack of interpretability of the machine learning methods, especially the neural models [12]. Getting the top features from a trained method allows us to understand and make sense of the information these methods process in the auto-separation of data into suicidal versus non-suicidal classes.

The limited number of available benchmarks is a challenge in suicide research [12]. Collecting data, especially under clinical supervision, is difficult and expensive. Meanwhile, social media data with actual ground-truth availability are limited, share-restricted, and come with other challenges. Community efforts are dedicated to attacking these challenges like working on small gold-standard, share-restricted data [13]. Due to these challenges, studies are limited to work on a single, limited collection and report only within-dataset evaluations. One drawbacks of this limitation is how the trained models would generalize to other data in a real-life setting remains unknown. Also, due to the lack of generalization experiments, researchers cannot determine whether their methods learn patterns or clues related to suicide. Ribeiro et al. shows how the deep learning models can ”learn” false features while returning high scores on within-corpus evaluations [25]. Evaluating the trained models on other data would allow researchers to realize whether the models have been overfitting to the training dataset.

3 Materials and Methods

3.1 Data

Each of the three collections comprises transcribed interviews conducted by clinical experts asking five “ubiquitous questions” to the volunteering patients. These questions were developed to initiate a conversation to harvest language from them: “Do you have hope, fear, secrets? Are you angry? Does it hurt emotionally?” [26]. The patient responses are long, containing more than one sentence, and are of free form. In this study, only the patient responses are used. These three collection studies have been approved by the Cincinnati Children’s Hospital Medical Center’s (CCHMC) Institutional Review Board.

Table 1: For each corpus used in the study, the average age of the patients, and the number of participants in each class.

Dataset	Average age	Suicide	Control
First corpus	15.5 ($\pm$ 1.4)	31	30
Second corpus	33.5 ($\pm$ 16.4)	130	123
Third corpus	19.2 ($\pm$ 3.0)	32	77

The first corpus in Table 1 contains data from patients in the Emergency Department admitted for suicide-related events, and the control set comprises orthopedic injury patients with no recent or past suicidal ideation [27]. The second corpus is similar to the first, except it was obtained from three hospitals in two cities:

CCHMC, University of Cincinnati Medical Center, and Princeton Community Hospital [21]. This corpus also includes follow-up interviews conducted a month after discharge. A recent longitudinal study found no significant difference in language a month after discharge [15], so the follow-up responses are also included in the corpus. Since our subject area is the detection of suicidal ideation, mentally ill patient interviews are excluded from the collection. Next, these two collections are merged into a single corpus to increase the training set size, containing 470 texts where the number of suicidal and control texts are balanced.

The third corpus is different from the first two as the participants are adolescent epilepsy patients of CCHMC. The original aim of this collection was to identify psychiatric comorbidities in these patients [28]. The number of suicidal and control subjects in this collection are in Table 1. The ideation levels for the suicidal group are lower than those of the previous two corpora like in a real-life population. Most of the suicidal group patients admit having suicidal thoughts in the past, which does not mean they are currently suffering from suicidal ideation, similar to real-life, i.e. fewer suicidal individuals, varying ideation levels [29]. It also includes some follow-up interviews, overall comprising 213 texts.

3.2 Statistical Features

We apply the standard text pre-processing methods (lowercase conversion, punctuation removal - prohibited for bigrams, and tokenization) before extracting features. The first feature set has unigrams, i.e. individual word frequencies. We exclude words occurring in fewer than five training set texts ($< 1.3\%$ of the set) to eliminate misspellings or infrequent words. We also experiment with stopwords to evaluate their effect as there is no consensus on removing or keeping them. For ease, we use the stopwords list in the NLTK library and use them as unigrams [30]. Subsequently, we experiment with bigram features that are concurrently occurring word pairs. For their extraction, we also employ the NLTK library [30]. The end-of-sentence punctuations are kept during pre-processing before bigram extraction to include the information on which words are at the beginning or end of a sentence. We used the same threshold as before (< 5) to remove the infrequent bigrams. Finally, we obtain the n-grams features (n = $\{1, 2\}$), which are the most popular textual features in clinical studies ([15, 21, 31]) by merging the unigram and bigram features. After the statistical feature extraction, we transform the feature values into their natural logarithms to decrease differences in magnitude, especially in n-gram features where frequent words might bias the performance. Finally, we normalize each vector to unit length.

3.3 Lexical Associative Networks

One way to determine associations from textual data is to obtain joint usage information, i.e., two words that are frequently used together in the same textual neighborhood are considered to be more related [32, 33, 34, 35]. Building networks from a training corpus is a way of condensing the word relations, ideally capturing the shared thoughts and ideas they represent. Connecting the words or tokens in a text corpus based on such associations results in a lexical associative network. Two sub-sets of suicidal and control, undirected and weighted networks created from word associations are present in Figure 1 where the weights are computed as the correlation coefficient:

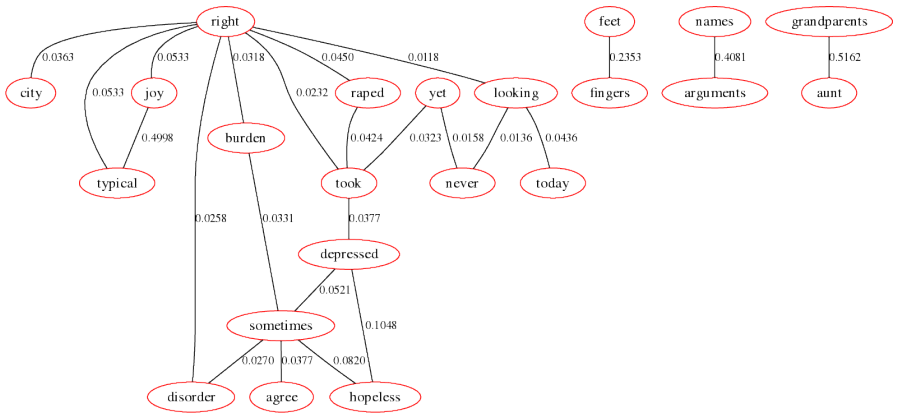
$$w_{i,j} = \frac{p(i,j) - p_i p_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}}. \tag{1}$$

p(i,j) is the probability of two words i and j to co-exist in the same sentence, while p_i and p_j are the individual occurrence probabilities.

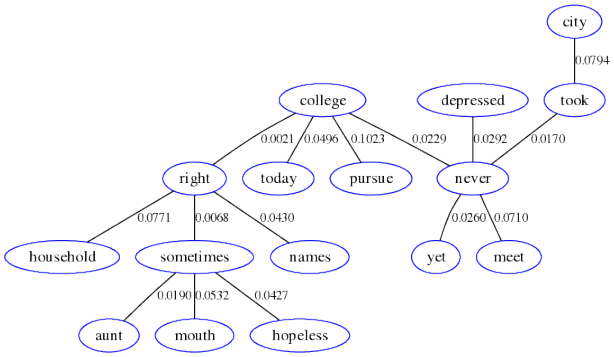
One way of using these networks is for text classification. A method named excess weight density (EWD) has been proposed a semi-supervised approach [22, 36]. It estimates how well the words in a given text fit either network using the connection weights and the network density computation. For classification, it assigns the label of the network with the highest connection density to the text. We use this method as a baseline in this study.

input : $G(V, W^1)$ and $G(V, W^2)$; // Lexical networks of two classes
output: $\vec{x}, \vec{l}$; // Feature vector, label vector
$\vec{x}$ =[ ] ; //
$\vec{l}$ = [ ] ; //
for $\forall (i,j) \in V$; // For all node (word) pairs
do
  if $(w^1_{i,j} > \lambda \in W^1) \vee (w^2_{i,j} > \lambda \in W^2)$; // λ threshold
  then
    $x_{i,j} = w^1_{i,j} - w^2_{i,j}$; // Compute the feature value
    $\vec{x}$ += $[x_{i,j}]$; // Add the feature
    $\vec{l}$ += $[(i,j)]$; // Add the labels
  end
end

Algorithm 1: Algorithm for extracting lexical associative network features.



(a) A sample suicide network



(b) A sample control network

Figure 1: Randomly selected subsets from the suicide and control networks constructed using the associations between the words, where numbers on the connections are association weights.

Alternative to the text classification, these networks can also be used for feature extraction so they can be used together in machine learning tasks. For two networks $G(V, W^1)$ and $G(V, W^2)$ with the same nodes V, Algorithm 1 shows how the network features are extracted. For two words that exist in the networks as nodes (i, j), their connection weights are subtracted from one another to obtain the final feature value. A positive feature value means the words are strongly connected in the first (suicidal) network compared to the other (control), and vice versa for the negative values.

3.4 Supervised Classifiers

In the experiments, three supervised methods are used. The first method is a logistic softmax. Logistic regression is popular in clinical studies [37, 31, 11], and the logistic classifier constructed as a neural network with no hidden layers is equivalent to logistic regression when the softmax is used in the output layer. Using the Tensorflow API [38], we create this no hidden layer network and apply standard feed-forward back-propagation for training. To increase the complexity, a multilayer perceptron (MLP) is created by adding a hidden layer with a thousand neurons and hyperbolic tangent as the activation function, and a dropout rate of 0.98. Finally, to further increase the model complexity, a Convolutional Neural Network (CNN) is selected for its popularity in suicide research [17, 18, 12]. We use Kim's approach for sentence classification [39], which contains five layers (an input layer converting words into numbers representing their location in the vocabulary, an embedding layer that learns the semantic relations between the words, a convolution layer that applies convolution filters of different window sizes on embedding vectors, a max-pooling layer where the maximum value - assumed to be the most important feature - is extracted and returns "important" dense features, and finally, the softmax layer makes a classification decision). We modify this model to learn and classify complete texts instead of sentences. We initialize the embedding layer with pre-trained word2vec vectors that were trained by the Google News corpus, containing 100 billion words[1]. We keep the remaining parameters at default: the dropout rate of 0.5, the decay rate of 2.5, the batch size of 64, the number of filters as 128, and the filter sizes $\{3, 4, 5\}$. To avoid overfitting, we use early stopping with the patience of 25 epochs on the validation set during training.

4 Results

For the experiments, from the merged two collections containing 470 interviews, we randomly select 25 suicidal and 25 control texts as the excluded test set. Next, we divide the remaining 420 texts into training and validation sets repeatedly ten times (ten-fold) at random as part of a Monte Carlo cross-validation to ascertain different texts become training set at each fold. Followingly, we perform the following operations at each fold: we randomly select 60 suicidal and 60 control texts for validation, and the remaining becomes the training set, and subsequently, we use the validation sets to tune the models during training and then classify the excluded test set and the third generalization corpus. At the end of the ten folds, we average the classification scores. Since the validation set is used for parameter tuning. To make sure there is no overfitting, validation set classification results are

[1] https://code.google.com/archive/p/word2vec/

compared to those of the test sets. Later, the validation sets are fully excluded from the remaining experiments.

Table 2: Average AUC scores and the standard deviation over ten-fold within-corpus and generalization evaluations.

Test set (within-corpus) results

Methods	unigrams	stopwords	unigrams without stopwords	bigrams	n-grams ($n \leq 2$)	lexical network features
Logistic	92.0 (2.0)	74.8 (5.7)	92.5 (1.0)	87.8 (1.9)	90.7 (3.0)	85.5 (2.1)
MLP	95.1 (1.2)	77.9 (3.7)	94.4 (1.1)	89.3 (1.8)	92.4 (1.8)	86.1 (2.9)

Third (generalization) set results

Methods	unigrams	stopwords	unigrams without stopwords	bigrams	n-grams ($n \leq 2$)	lexical network features
Logistic	66.1 (1.2)	59.3 (2.0)	66.5 (1.8)	68.8 (1.4)	68.5 (1.4)	62.5 (2.0)
MLP	65.8 (1.8)	58.7 (1.5)	65.3 (1.5)	68.5 (1.5)	68.7 (1.4)	62.0 (1.9)

Table 3: Average AUC scores and the standard deviation over ten-fold within-corpus and generalization evaluations of the methods not trained by statistical features.

Test set (within-corpus) results

Methods	Results
EWD	77.6 (3.1)
CNN	88.5 (2.2)

Third (generalization) set results

Methods	Results
EWD	64.3 (2.1)
CNN	61.6 (2.5)

Table 2 and Table 3 illustrate the within-corpus evaluation results and the generalization results as average and standard deviation of AUC values on the test set and on third set. The following items are the main observations we captured from these results:

- Results show that machine learning methods trained by lexical associative network features provide high performance, but fall short compared to the

models trained by standard features (unigrams, bigrams, n-grams). When used together with n-gram features to increase the heterogeneity of the feature space, these features may enhance the performance.

- The comparison of machine learning methods shows that neural networks and logistic classifier provide high performance independent of the feature type on within-corpus or generalization experiments. Also, CNN performs high on within-corpus experiments but fails to generalize over the third corpus as effectively as the MLP, or the EWD. Overall, no single method outperforms the rest under all circumstances, but the MLP seems to be the best.

- A notable observation from the results is that the simple, no hidden layer neural network, which is the logistic classifier, has performance comparable to and often not far behind the CNN and the MLP. This observation is significant for clinical applications because logistic classifier is explainable and computationally simple. These results also indicate that complex approaches are driven more by historical practice than tangible benefits. The superiority of simple methods over the deep learning methods on small suicide corpora is commonly observed [14].

In addition to the classification experiments, we also take advantage of the fact that some machine learning methods are not black-boxes. The learned weights of the trained logistic classifier can be used to determine which words were most important for returning the above AUC scores. Thus, we use the trained logistic classifier and return top ten features for each feature set in Table 4. The top unigrams contain suicidal words like "depression," "feeling," and "pill". Meantime, the top suicidal stopwords are "you" and "we." Top bigrams and n-grams also have word pairs associated to depression, anger, feeling, while the control features have happiness and laughter-related words. Overall, the statistical features are interpretable. Yet, interpreting the top lexical associative network features is not as straightforward due to their computational nature. For example, the word pair "hope-role", unlike bigrams, does not mean it was more frequent in the suicidal set. It means, its feature value, which could be negative, was more significant for identifying the suicidal interview transcripts. Thus, the negative correlations are also in effect in these network features.

5 Discussion

Perhaps the most interesting observation from the results is the performance of the extremely simple logistic classifier. Over the years, researchers have applied increasingly sophisticated machine learning methods to text classification - often with

Table 4: Top ten features from to the trained logistic classifier of the remaining feature sets (S: top suicidal, C: top control class features.)

Feature	Top ten features per class (left to right: most to less important)
unigrams	S: depression, feeling, feels, medication, depressed, thoughts, kinda, pills, met, working C: role, laughs, play, healthy, say, fine, something, sports, upset, passed
stopwords	S: than, or, does, if, you, can, we, further, you've, over C: here, hasn't, which, against, how, this, herself, himself, was, did
bigrams	S: i'm-angry, depressed-and, yeah-it, it-feels, because-of, feels-like, i'm-at, feel-like, i-need, this-to C: laughs-i, .-laughs, me-angry, big-role, role-., close-., no-it, plays-a, happy-person, i-like
n-grams	S: feeling, depression, working, thoughts, i'm-angry, yeah-it, because-of, depressed-and, medication, pills C: laughs-i, role, .-laughs, laughs, play, big-role, role-., me-angry, close-., they-say
lexical network features	S:role-really, hope-role, role-secrets, life-role, role-um, things-role, people-role, anything-stupid, much-role, better-role C: laughs-god, think-depression, feeling-um, good-point, feeling-that's, something-good, life-something, takes-see, get-working, laughs-secrets

good results. However, the experiments show that its performance is competitive with more complex machine learning methods such as MLP, though the latter did show slightly better average AUC. More notably, the logistic classifier produces the best generalization performance across the board. The generalization results also showed that different features (except stopwords only) made relatively little difference. Together, these observations indicate that the essential information used for inference by the machine learning classifiers is relatively simple and is almost entirely captured by simple (unigram) features and a simple (logistic) classifier. In addition to the computational advantage conferred by these simpler models, their use also leads to greater explainability. When unigrams are used as features in a non-hidden-layer logistic classifier, it is very straightforward to determine the differential value of words (or word combinations if n-grams are used) through feature analysis as demonstrated earlier.

The experiments on stopwords indicate that: a) machine learning classifiers can

achieve performance using only stopwords as features, suggesting that the stopword distribution alone does provide some information about the suicidality of a text; b) When machine learning classifiers are trained by unigrams, including or excluding stopwords makes no significant difference, suggesting that whatever information is carried by stopwords is also available in the unigrams (see Table 4). One can conclude from (b) that choosing either approach is accurate. Since classifying the stopwords alone shows some benefit, it might be beneficial to include them. Meanwhile, the lexical associative network features alone failed to overpower the performance of the statistical features, which are powerful. Yet, these network features alone provided comparable performance to the well-established statistical features. Especially in the generalization experiment, the lexical network features trained by either machine learning method provides better performance than CNN. Also, the network-based classifier EWD also performs better than CNN, and is comparable to the other machine learning methods in the generalization experiment. These findings show that such networks deserve further studying and improvement.

A key point of this study is that it provides a comparison between two different clinical datasets. Most studies use clinical collections that cannot be released for legal reasons. This limitation eliminates the possibility of other researchers investigating methods on the same corpus. Nevertheless, this can be overcome to some extent by testing methods developed on one corpus on another available collection, giving a sense of how transferable such models can be. It also illustrates the classifiability of the collection used for training compared to other studies. Classification results reaching as high as AUC=$95 \pm 1.2\%$ on the test set confirm the previous literature reporting high within-corpus results from these collections [21, 27, 15]. Meanwhile, the generalization of the trained models on the third collection returns around AUC=$68.8 \pm 1.4\%$. Although it may seem low compared to the within-collection results, it is necessary to remember that the methods were not trained on this collection. The contents of this corpus are different from the previous two datasets as the reported suicidal ideation levels are low or were in the past. A previous study reports an AUC score of 71% on within-corpus evaluation using n-grams and SVM on this third set [28]. The trivial difference between 68.8 and 71 confirms the low levels of suicidal ideation. The fact that the models trained on different datasets could classify this distinct collection almost as good as its past within-dataset results is a rich contribution for the future of suicide risk detection applications. This outcome proves that the identifiers of suicidal ideation present within the features are transferable to other collections. These results are especially significant considering the size of the training data since working with machine learning methods on small datasets is a challenge [14]. Overall, these observations confirm that trained machine learning methods and the networks together have a great potential for use in future mental health datasets.

6 Conclusion

Implementing machine learning methods is becoming a common practice in suicide research. Yet, the literature needs more experiments for enhancing the confidence towards these research studies. The current state of research is far from perfection for the lack of clinically-labeled data, lack of generalization experiments, and lack of different method evaluations. This study responds to these issues by experimenting with statistical text features, and constructing networks for feature extraction and classification. Among the results of many folds, machine learning methods prove to be successful in detecting suicidal ideation with an average AUC of 95% on within-corpus evaluations on the merged two collections. Most essentially, the trained models achieve success in classifying the low ideation levels in the third collection. Tests on lexical networks also show promise as a classifier and as a source of features. Meantime, logistic classifier performs almost as well as, and at times even better than complex, deep learning methods, and in addition, promises explicability through feature analysis. However, extensive future work with more methods and more data is needed before clinicians start utilizing these models in practice. Until then, these experiments help improve confidence in employing them so new studies can explore these aspects further and enhance the state of knowledge.

References

[1] Centers for Disease Control and Prevention. Preventing Suicide. U.S. Department of Health & Human Services; 2018. Available from: https://www.cdc.gov/violenceprevention/pdf/suicide-factsheet.pdf.

[2] World Health Organization G. Suicide: Key Facts. World Health Organization; 2019. Available from: https://www.who.int/news-room/fact-sheets/detail/suicide.

[3] Sher L. The impact of the COVID-19 pandemic on suicide rates. QJM: An International Journal of Medicine. 2020;113(10):707–712.

[4] Mamun MA, Griffiths MD. First COVID-19 suicide case in Bangladesh due to fear of COVID-19 and xenophobia: Possible suicide prevention strategies. Asian journal of psychiatry. 2020;51:102073.

[5] Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. Journal of medical Internet research. 2020;22(10):e22635.

[6] Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, et al. Prevalence and treatment of mental disorders, 1990 to 2003. New England Journal of Medicine. 2005;352(24):2515–2523.

[7] Kazdin AE. Evidence-based treatment research: Advances, limitations, and next steps. American Psychologist. 2011;66(8):685.

[8] Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomedical informatics insights. 2018;10:1178222618792860.

[9] Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychological Bulletin. 2017;143(2):187.

[10] Murray D. Is it time to abandon suicide risk assessment? British Journal of Psychiatry Open. 2016;2(1):e1–e2.

[11] Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. Psychological science. 2010;21(4):511–517.

[12] Ji S, Pan S, Li X, Cambria E, Long G, Huang Z. Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems. 2020.

[13] Macavaney S, Mittu A, Coppersmith G, Leintz J, Resnik P. Community-level research on suicidality prediction in a secure environment: Overview of the CLPsych 2021 shared task. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; 2021. p. 70–80.

[14] Bayram U, Benhiba L. Determining a Person's Suicide Risk by Voting on the Short-Term History of Tweets for the CLPsych 2021 Shared Task. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; 2021. p. 81–86.

[15] Pestian J, Santel D, Sorter M, Bayram U, Connolly B, Glauser T, et al. A Machine Learning Approach to Identifying Changes in Suicidal Language. Suicide and Life-Threatening Behavior. 2020.

[16] Resnik P, Foreman A, Kuchuk M, Musacchio Schafer K, Pinkham B. Naturally occurring language as a source of evidence in suicide prevention. Suicide and Life-Threatening Behavior. 2020.

[17] Tadesse MM, Lin H, Xu B, Yang L. Detection of suicide ideation in social media forums using deep learning. Algorithms. 2020;13(1):7.

[18] Gaur M, Alambo A, Sain JP, Kursuncu U, Thirunarayan K, Kavuluru R, et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In: The World Wide Web Conference; 2019. p. 514–525.

[19] Sawhney R, Manchanda P, Mathur P, Shah R, Singh R. Exploring and learning suicidal ideation connotations on social media with deep learning. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2018. p. 167–175.

[20] Matero M, Idnani A, Son Y, Giorgi S, Vu H, Zamani M, et al. Suicide risk assessment with multi-level dual-context language and bert. In: Proceedings of the sixth workshop on computational linguistics and clinical psychology; 2019. p. 39–44.

[21] Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide and life-threatening behavior. 2017;47(1):112–121.

[22] Bayram U, Minai AA, Pestian J. A Lexical Network Approach for Identifying Suicidal Ideation in Clinical Interview Transcripts. In: International Conference on Complex Systems. Springer; 2018. p. 165–172.

[23] Teixeira AS, Talaga S, Swanson TJ, Stella M. Revealing semantic and emotional structure of suicide notes with cognitive network science. Scientific reports. 2021;11(1):1–15.

[24] De Beurs D, Fried EI, Wetherall K, Cleare S, O'Connor DB, Ferguson E, et al. Exploring the psychology of suicidal ideation: A theory driven network analysis. Behaviour research and therapy. 2019;120:103419.

[25] Ribeiro MT, Singh S, Guestrin C. Why should i trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016. p. 1135–1144.

[26] Pestian JP, Matykiewicz P, Duch W, Glauser TA, Kowatch RA, Grupp-Phelan JM, et al.. Processing text with domain-specific spreading activation methods. Google Patents; 2016. US Patent 9,477,655.

[27] Pestian JP, Grupp-Phelan J, Bretonnel Cohen K, Meyers G, Richey LA, Matykiewicz P, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide and life-threatening behavior. 2015.

[28] Glauser T, Santel D, DelBello M, Faist R, Toon T, Clark P, et al. Identifying Epilepsy Psychiatric Comorbidities With Machine Learning. Acta Neurologica Scandinavica. 2019.

[29] Gibbons RD. The statistics of suicide. Shanghai archives of psychiatry. 2013;25(2):124.

[30] Bird S, Klein E, Loper E. Natural Language Processing with Python. O'Reilly Media; 2009.

[31] De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM; 2016. p. 2098–2110.

[32] Marinho VQ, Hirst G, Amancio DR. Authorship attribution via network motifs identification. In: Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on. IEEE; 2016. p. 355–360.

[33] Mehri A, Darooneh AH, Shariati A. The complex networks approach for authorship attribution of books. Physica A: Statistical Mechanics and its Applications. 2012;391(7):2429–2437.

[34] Lahiri S. Complexity of word collocation networks: A preliminary structural analysis. arXiv preprint arXiv:13105111. 2013.

[35] Bayram U, Roy R, Assalil A, BenHiba L. The unknown knowns: a graph-based approach for temporal COVID-19 literature mining. Online Information Review. 2021.

[36] Bayram U, Pestian J, Santel D, Minai AA. What's in a Word? Detecting Partisan Affiliation from Word Use in Congressional Speeches. In: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE; 2019. p. 1–8.

[37] Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science. 2017;5(3):457–469.

[38] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. Software available from tensorflow.org. Available from: https://www.tensorflow.org/.

[39] Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:14085882. 2014.