Suicide Ideation Estimators within Canadian Provinces Using Machine Learning Tools on Social Media Text

Ruba Skaik and Diana Inkpen
School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada
Email: {rskai034, Diana.Inkpen}@uottawa.ca

***Abstract*—Suicide has become one of the leading causes of death worldwide. It is a serious public health problem, and the right prompt response can mitigate it. Therefore, identifying individuals with suicide risk and offering immediate counseling to everyone that might need it is a crucial step. In this research, we utilize personal narratives collected through the popular social media website (Reddit) to build a model suitable for predicting suicide ideation in a sample of Twitter users that is representative for the Canadian population. The labeled dataset contains only 621 users, and with that limited number of training instances we extracted features for classical machine learning and achieved an F1-score of 0.922 using linguistic and emotion features. In addition, we fine-tuned a Sentence Pair Classification BERT model and achieved 92.6 F1-score. The classical machine learning trained model was applied on Canadian population representative dataset. The geographic and demographic patterns of suicide ideation correlate with the suicide statistics reported by Statistics Canada for 2015.**

***Index Terms*—natural language processing, suicide, mental health, social media, deep learning**

I. Introduction

Suicide is a major public health problem and a growing risk worldwide. Canada's youth suicide rate is the third highest in the industrialized world. In 2015 there were 4,405 suicides in Canada, a rate of 11.5 per 100,000 people as stated by Statistics Canada. Each suicide case has significant consequences on the physical and emotional well-being of families and societies in general. Early detection of Suicide Ideation (SI) is a significant step in preventing suicide. Social media is progressively regarded as a valuable source for public health monitoring. Among social media websites, Reddit has emerged as a widely used online message board for focused mental health topics including depression, addiction, and suicide. In particular, the SuicideWatch subreddit has more than 170,000 subscribers who can post anonymously about what they are going through and receive emotional support from other users. On the other hand, Twitter is the third most popular social network website in Canada and had over 7.5 million users in 2017[1]. Twitter is a rich environment where people share their thoughts, life-events, and feelings throughout the day. In this research, we are combining the power of Reddit and Twitter environments to predict Suicide Ideation (SI) at population level using only the textual information expressed by the users.

II. Related Work

Social media has attracted many researchers to analyze its contents for the benefit of humanity [1]. Current research reveals significant correlations between the mental health of a person and the linguistic material they post on social media. For instance, O'Dea *et al.* [2] used Support Vector Machine (SVM) algorithm with TF-IDF features to distinguish strongly concerning suicide-related tweets among 14,701 tweets with an accuracy of 80%. Wang *et al.* [3] performed the same on Chinese online communities. In [4], Braithwaite *et al.* used decision tree classification on Linguistic Inquiry and Word Count (LIWC) of Twitter data from 135 participants to measure suicide risk at user-level with an accuracy of 92%. Nobles *et al.* [5] fed psychology linguistic features into a deep neural network (MPL) to predict periods of suicidality using personal communication and social media data. Using Reddit data, Grant *et al.* [6] used k-means clustering to extract latent topics of SI. Similarly, Aladag *et al.* [7] used SVM and logistic regression (LR) on the TF-IDF, LIWC and sentiment analysis features of Reddit data to predict SI.

Having a comprehensive and accurate dataset is a critical success factor for creating reliable machine learning models. Wang *et al.* [8] randomly crawled one million users (394 million postings) and used a keyword-based method to identify users at risk of suicide and used linguistic analysis to explore behavioral and demographic characteristics. Whereas, Lv *et al.* [9] used Sina Weibo to build a suicide dictionary that significantly correlated with expert ratings in detecting suicidal expression and predicting users at risk.

Predicting SI from social media is considered one step towards identifying affected groups based on gender, age, geographic location, or other characteristics to pay more attention to these groups and to offer psychological

Manuscript received December 5, 2020; revised June 19, 2021.

[1] https://www.statista.com/topics/2729/social-networking-in-canada/

doi: 10.12720/jait.12.4.357-362

services. The association between suicidal ideation and linguistic features was examined [10], [11]. In [12], Vioulès *et al*. proposed a framework for real-time detection of suicide-related tweets based on ensemble method lexicons. From the CLPsych 2019 shared task [13], the most related activity to our study is the binary classification for detecting "flagged" users who are at risk of suicide. The best performing system for this task, achieved 0.922 F1-score for differentiating at-risk and non-suicidal users within the SucideWatch subreddit and 0.843 F1-score among all subreddits [14]. They used GloVe and ELMo word embeddings as an input for the following neural networks (CNN, Bi-RNN, Bi-LSTM or Bi-GRU), then applied three types of pooling to be fed to the fusion component that contains the neural features and the classes predicted probability distribution. At the end of the pipeline, an SVM classifier was used to classify at-risk users. De Choudhury *et al.* [15] built a statistical approach on data from Reddit users who shifted from mental health concerns to SI. Their approach derives markers to detect this transition incident through the three cognitive psychological integrative model of suicide including thinking, ambivalence, and decision making. Benton *et al.* [16] presented a Multi-task Learning (MTL) model using a feed-forward network with a character n-gram input layer to predict potential suicide attempts and the presence of atypical mental health. Multiple studies have investigated the factors correlated with completed suicide [17].

III. SI Classification Models

We trained our models on a Twitter dataset denoted as $\mathcal{S}2$ and tested them on Reddit dataset denoted as $\mathcal{S}1$. The datasets are described in Section IV. Before extracting the features from the text, we performed several pre-processing steps to reduce the noise from the original data and we concatenated all the posts of each user into a single document to represent the user. Then we evaluated four classical machine learning models and four deep learning models to identify at-risk users based on their posts/tweets. Accuracy, precision, recall and the F1-score were calculated to assess the performance of each classifier using 5-fold cross-validation with stratified sampling.

A. Classical Classifiers

We used the following classical classifiers: SVM, LR, Random Forest (RF), and Extreme Gradient Boosting (XGBoost), using statistical, linguistic, emotional and topic modeling features as shown in Table I. The statistical features include the document length, word counts, number of punctuation marks, emojis and emoticons used, frequency of posting, as well as TF-IDF values for unigram and bigrams. TF-IDF represents the significance of a term in a document based on the frequency of its appearance in the document, and inversely proportional to its frequency in the rest of the corpus. The initial number of features was 7,637 and 130,517 for uni-grams and bi-grams, respectively. Then, we used Principal Component Analysis (PCA) for dimensionality reduction that reduced the TF-IDF features to 325 features. Gensim [2] Latent Dirichlet Allocation (LDA) topic modeling tool was used to extract L hidden topics based on a coherence test. The resultant L topic distributions were fed into the classifiers as L features. In our case, they give the highest coherence value when L is set to 20 topics. Finally, linguistic features were obtained using LIWC [3] dictionary. LIWC defines the language patterns of the posts and categorizes them in psychologically meaningful groups. Also, emotion features were estimated using VADER (Valence Aware Dictionary and sEntiment Reasoner) tool. VADER is a rule-base model used for social media text sentiment analysis. VADER estimates the sum of all the lexicon ratings and assigns a probability for negative, positive or neutral sentiments, and it resembles the annotations of human raters with 0.88 correlation coefficient [18]. In addition, we extracted all the emoticons and emojis from the original post and replaced them with their meaning.

TABLE I. Summary of the Features Used for Predictive Models for SI on the Dsataset $\mathcal{S}1$ with the Best Results Achieved

Feature Set	Classifier	Acc.	Prec.	Recall	F1-score
Stat+tfidf	SVM	80.8%	0.844	0.911	0.875
	LR	80.6%	0.844	0.907	0.874
	RF	77.6%	0.793	0.946	0.862
	XGBoost	78.7%	0.822	0.912	0.864
LIWC	SVM	87.1%	**0.907**	0.920	0.913
	LR	87.2%	0.904	0.927	0.915
	RF	86.7%	0.886	0.941	0.913
	XGBoost	86.1%	0.894	0.924	0.908
LDA	SVM	80.5%	0.837	0.917	0.874
	LR	80.6%	0.839	0.917	0.875
	RF	77.6%	0.790	**0.951**	0.862
	XGBoost	75.8%	0.812	0.882	0.844
Emo+LIWC	SVM	84.5%	0.896	0.896	0.895
	LR	86.6%	0.900	0.922	0.911
	RF	87.0%	0.886	0.948	0.915
	XGBoost	**88.1%**	0.903	0.942	**0.922**

B. Deep Learning Models

Deep learning models take the word-embedding matrix as the input of each model, the vocabulary contains 119,626 unique words, from which we selected the most frequent 50,000 words. The output is a single sigmoid activation function. The layers in between are sequential and fully connected; a dropout layer of probability 0.3 is used to avoid over-fitting. For the base model, we trained a 300-dimensional word embedding layer on the SuicideWatch subreddit posts then passed it to a bidirectional Gated Recurrent Unit layers (GRU). We use global max pooling with ReLu activation layer. And since we have a small dataset, we employed transfer learning model to enrich the training models as follows:

- **BiLSTM_FastText:** We used the pretrained 300-dimensional FastText based on Common Crawl [4] as the word embedding layer, followed by a

[2] https://radimrehurek.com/gensim/models/ldamodel.html
[3] https://liwc.wpengine.com
[4] https://fasttext.cc

bidirectional Long Short-Term Memory (LSTM). We used max pooling and average pooling to select the most representative features.

- **CNN_FastText_LIWC:** Inspired by [19] we applied 2D convolutions and 2D max pooling instead of 1D max pooling to sample more meaningful information from FastText word embedding layer. The convolution layer has 32 filters with the sizes of 2, 3 and 5. LIWC features were used in conjunction with the concatenated output of the max pooled word embedding convolution layers. Finally, the decision is made by sigmoid function as shown in Fig. 1.
- **BertForSequenceClassification:** Lately, BERT became the most popular NLP approach to transfer learning [20]. Using Huggingface[5] abstraction, we fine-turned the pre-trained BERT model using "Sentence Pair Classification" model (BertForSequenceClassification) to obtain an F1-score 0.926 and an accuracy of 89.4% as shown in Table II.

TABLE II. DEEP LEARNING MODELS ON $\mathcal{S}1$ DATASET

Model	Acc.	Prec.	Recall	F1-score
Bi-GRU(Baseline)	85.6%	0.869	0.954	0.908
Bi-LSTM+FastText	85.8%	0.877	0.943	0.907
CNN_FastText+LIWC	87.0%	0.879	**0.959**	0.915
BertForSequenceClassification	**89.4%**	**0.900**	0.955	**0.926**

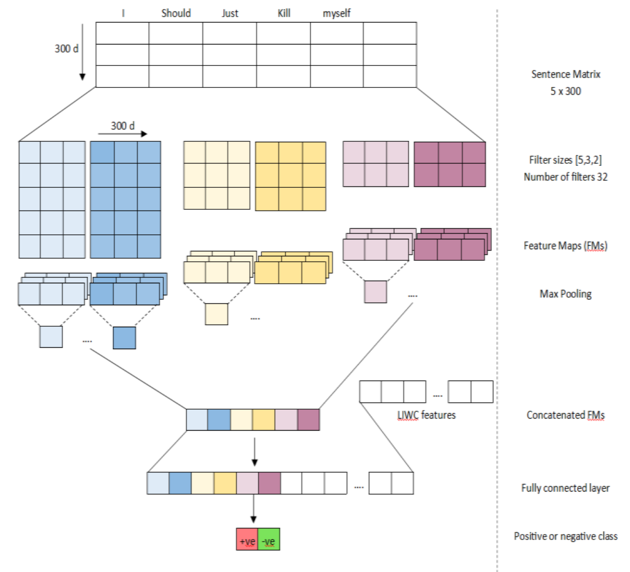

Figure 1. CNN_FastText_LIWC architecture.

IV. DATASETS

A. Datasets Overview

For population suicide prediction, we conducted our experiments on three datasets: Two labeled datasets for SI $\mathcal{S}1$ and $\mathcal{S}2$, and one dataset for Canadian population inference $\mathcal{P}1$ as summarized in Table III.

TABLE III. SUMMARY OF THE DATASETS USED IN THIS STUDY

Ref.	Platform	Notes	Users	Posts
$\mathcal{S}1$	Reddit	Task (A)	621	1,105
$\mathcal{S}2$	Twitter	SI	495	987,870
$\mathcal{P}1$	Twitter	Canadian	278,627	10,594,841

B. CLPsych Dataset

The $\mathcal{S}1$ dataset is based on the CLPsych 2019 shared task (A). It consists of the posts of users who posted only on the SuicideWatch subreddit. The severity of the users' status was annotated by crowdsourcers into four classes: non-suicidal, low-risk, high-risk and severe, with Krippendorff's α= 0.55. The $\mathcal{S}1$ dataset contains 462 users with suicide thoughts and 159 non-at-risk users. The $\mathcal{S}1$ dataset only contains the posts written by the users including the subreddit name, the title of the post and the full text along with the time of posting. The data does not contain any profile information. For the purpose of this research, we considered the distinction between safe and at-risk users, regardless of the degree of risk.

C. Twitter Datasets

The $\mathcal{S}2$ and $\mathcal{P}1$ are subsets of the Twitter dataset that was collected by Advanced Symbolics Inc. (ASI)[6]. By 2018, they have collected millions of tweets for 278,627 users using the Conditional Independence Coupler (CIC) sampling algorithm that is based on Coupling from the Past (CFTP) with enhancing the stopping condition by measuring how far is the chosen node (user) from the starting node on a smaller subset of the online network. The method adjusts the weights of sampling using poststratification to compensate for underrepresented groups of the population [21], [22]. $\mathcal{S}2$ dataset contains 495 users where 4 of them were labeled as at-risk by a specialized psychologist, whereas $\mathcal{P}1$ dataset contains population-representative Canadian users who tweeted during 2015 and it contains the following demographic information per user as described in [23]:

- **Spatial information:** The location of the Twitter users can be inferred either by using the GPS coordinates of the tweets - if enabled by the user from his/her mobile device - or by the self-declared location - set by the user in his/her profile page. If the location property is enabled, then the longitude and latitude points are fetched from the user's mobile phone GPS and stored with each tweet till it is turned off. The location of the user is predicted using the k-means algorithm to cluster the GPS coordinates, and presumes that the cluster of the greatest number of points as the user's location. If no geotagged tweets exist, then Microsoft's Bing Maps is used to look for the address specified in the user's profile (if feasible); otherwise the inferred-location field is left empty. The CIC algorithm can limit the searching within users in a certain geographic area.
- **Demographic information:** ASI also predicted users' demographics using signals from Twitter data and other related external resources. The age

[5] https://huggingface.co/transformers/model_doc/bert.html

[6] A market research company based in Ottawa.

and gender probability distribution are estimated by comparing the first name with Canada's birth records. The estimation is adjusted using life tables - that contains life expectancy and associated age and sex projections for Canada - then the profile photo was analyzed using Face++ API[7]. The age_gender probability distribution is deduced for each user in 12 fields as in (1). Let age $\mathbb{A}$ = {"<25","25-34","35-44","45-54","55-64",">65"} and gender $\mathbb{G} = \{\text{Male}, \text{Female}\}$.

$$Gender_{Age} = \{P(g_a): \forall g \in \mathbb{G} \wedge a \in \mathbb{A}\} \quad (1)$$

where $\sum_{(\forall g \in \mathbb{G})} P(g) = 1$ and $\sum_{(\forall a \in \mathbb{A})} P(a) = 1$.

The differences between the probabilities of each category vary. Thus, we decided to keep users with high confidence for both age and gender prediction based on the following rules: For gender, we assign to the user the gender of the maximum gender probability of all age groups with probability ≥92.5% as shown in (2):

$$\{max(P_M, P_F): |P_M - P_F| > \varepsilon\,;\ \varepsilon = 0.85\} \quad (2)$$

For age, we assign to the user the age group of the maximum age group probability P(α), given that the difference between the largest and the second largest P(β) is greater than ϵ, using the formula shown in (3):

$$P(\propto) \leftarrow max\{\textstyle\sum_{\forall \xi \in G} P(\propto) : \propto \in \mathbb{A}\} \quad (3)$$

where $P(\beta) \leftarrow max\{\sum_{\forall \xi \in G} P(\beta) : \beta \in \mathcal{B}\}$ and $B = \mathbb{A} - Age(P(\propto))$ and $|P(\propto) - P(\beta)| > \varepsilon$[8].

Demographic information is crucial and needs to be as accurate as possible. Therefore, we eliminated the records of the users who have age and gender prediction less than the defined confidence and users who do not have a well-defined location mapped to a Canadian province. Canada has ten provinces and three territories. Hence, the number of users reduced from 278,627 to 40,631. We compared the distribution of $\mathcal{P}1$ that has tweets posted during 2015, with the 2016 Canadian Census. The 2016 Canadian Census administered by Statistics Canada, is the most recent Canadian population estimate, with a total of 35,151,728 people. The difference is < 5% for each area, except for the province of Quebec as shown in Table IV and Table V. Quebec is the only province whose sole official language is French. Less than 10% of Quebecers are Anglophone. For this reason, and because the language of correspondence considered in this research is English, we exclude the province of Quebec from our experiments. In addition, Nunavut users are not present since the users' attributes failed to fulfil the previous confidence condition.

Regarding gender distribution in the provinces, the difference between the population share of the two genders in the 2016 census and in $\mathcal{P}1$ dataset is higher than 5% in the province of Saskatchewan and the Yukon territory. The female population is under-represented in both areas, and this needs to be considered when analysing our results. Other than that, all the other provinces and territories have a difference in the distribution between [0 - 5].

TABLE IV. $\mathcal{P}1$ DATASET AND CENSUS DATA FOR MALE(M) AND FEMALE(F) CANADIAN POPULATION BETWEEN 15 - 80 YEARS

Province	$\mathcal{P}1_M$	$\mathcal{P}1_F$	C_M	C_F
Newfoundland & Labrador	324	366	212,480	223,435
Prince Edward Island	223	231	56,715	60,430
Nova Scotia	1,015	1,060	371,380	396,745
New Brunswick	435	370	303,180	315,800
Quebec	2,070	1,611	3,272,335	3,370,080
Ontario	9,257	7,893	5,321,680	5,617,765
Manitoba	806	687	496,465	508,870
Saskatchewan	819	574	425,990	429,740
Alberta	2,831	2,480	1,616,775	1,607,865
British Columbia	3,776	3,533	1,881,805	1,965,670
Yukon	50	29	14,595	14,695
Northwest Territories	101	90	16,705	15,985
Canada	21,707	18,924	14,002,485	14,538,890

TABLE V. GEOGRAPHIC AND GENDER POPULATION DIFFERENCE BETWEEN CENSUS 2016 DATA AND THE $\mathcal{P}1$ DATASET

Province	D_M%	D_F%	$PR_{M/F}$	P-C%
Newfoundland & Labrador	0.05	0.12	1.79	0.17
Prince Edward Island	0.35	0.36	0.70	0.71
Nova Scotia	1.20	1.22	0.57	2.42
New Brunswick	0.01	0.20	5.06	0.19
Quebec	6.37	7.84	6.97	14.21
Ontario	4.14	0.26	5.33	3.88
Manitoba	0.24	0.09	4.60	0.15
Saskatchewan	0.52	0.09	9.01	0.43
Alberta	1.30	0.47	3.17	1.77
British Columbia	2.70	1.81	2.75	4.51
Yukon	0.07	0.02	13.47	0.09
Northwest Territories	0.19	0.17	1.77	0.36
Canada	4.36	4.36	4.36	0.08

V. RESULTS AND DISCUSSION

The combination of each set of features was experimented and showed the most important feature sets using different classical machine learning algorithms for flagged users' identification on the $\mathcal{S}1$ dataset. It contains the CLPSych 2019 training and test data combined. We did this to be able to test the generalizability of the model to run if on the population-level data. And, as mentioned earlier, we reported the results of the training dataset $\mathcal{S}2$ using 5-fold cross-validation. The association between emotion and linguistic features (LIWC) scored the highest accuracy (88.1%) and the best F1-score (0.922) using XGBoost algorithm, while Random Forest achieved the best recall score of 0.951 using topic modeling features.

Table II shows different architectures for deep learning models that achieved comparable and better F1-scores on the $\mathcal{S}1$ test dataset (0.926). The F1-scores that we obtained are comparable with the "(flagged) F1-score" of 0.922, the maximum F1-score achieved by the CLaC team in the CLPsych 2019 shared task [13].

VI. ANALYSIS

When restricted to limited training data and a small number of labeled users, using a good feature set and a robust classifier may lead to competitive results with

[7] A deep learning system developed by Megvii Technology to obtain face attributes including age and gender.
[8] We choose E=0.35, since each age group has a probability of 0.17.

deep learning algorithms. We showed that LIWC and emotions are the best discriminating features between the control group and the suicidal people. We tested our XGBoost model using LIWC and emotional features on randomly selected users from $\mathcal{S}2$ with an equal number of suicidal and control users, and the results showed an accuracy of 83.5% and F1-score of 0.791. Subsequently, we used the tuned model on the Canadian dataset $\mathcal{P}1$ for SI prediction. Based on Statistics Canada an estimate of the total number of people who reported serious suicidal thoughts is 3,396,700 persons based on data from 2015, i.e., 9.514% of the total population. Our model - for the same year - predicted 2,583 suicidal users (7.3%) on the sample population $\mathcal{P}1$. Fig. 2 and Table VI shows the distribution of age/gender demographics within the Canadian provinces. This adheres to the fact that males account for 75% of suicides among adults-based data from the Public Health Agency of Canada. shows the actual SI per 100,000 during 2015 for the 9 provinces according to Statistics Canada. The Pearson correlation between actual SI per-state and predicted user is 0.61, that indicates a reasonable correlation. Our predicted SI rates are comparable to the actual statistics as shown in Fig. 3, except for the province of Newfoundland and Labrador where it is underestimated, and Saskatchewan province where it is overestimated. The predicting model can be enhanced to handle multi-language analysis.

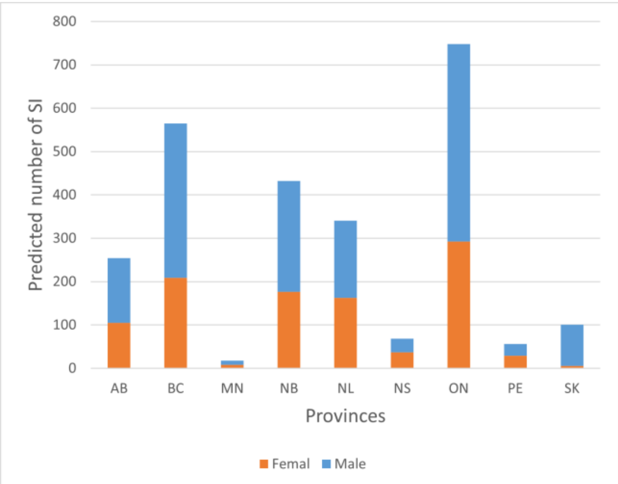

Figure 2. SI gender prediction.

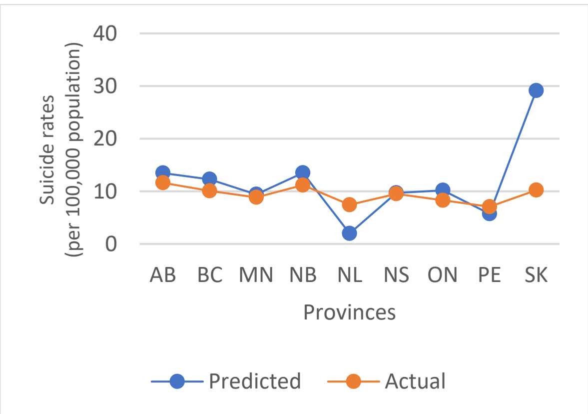

Figure 3. Predicted SI versus actual statistics.

TABLE VI. Actual and Predicted Suicidal thoughts per 100,000 Population in 2015 along with Age Predictions per Province

Pr.	>25	25-34	35-44	45-54	55-64	>65	Prd.	Act.
NL	25	45	55	106	104	6	13.46	11.62
PE	8	3	4	1	6	4	12.27	10.08
NS	5	18	12	19	22	2	9.42	8.85
NB	39	9	205	2	13	12	13.49	11.18
ON	88	62	103	253	175	77	1.99	7.43
MN	2	3	6	4	2	1	9.69	9.51
SK	71	9	9	21	75	16	10.19	8.28
AB	98	41	61	44	6	14	5.73	7.07
BC	73	81	107	194	115	47	29.13	10.22

VII. Conclusion and Future Work

Using machine learning for SI prediction at population level is feasible. The predictive power of machine learning algorithms can be utilized to assist health care providers and decision makers for better planning, community awareness, and medical care coverage. Our study showed the possibility of using classical machine learning tools to advance in this sensitive area. It is also worth mentioning that deep learning performs better without features engineering. Finally, adding engineered features to a Deep learning architecture like CNN may lead to a better performance. The model was trained on Reddit data and tested on Twitter data to check it generalizability, then applied to a population representative dataset and produced correlated results. In future work, we plan to consider multi-language analysis and to use the word embeddings with different settings and using hierarchical attention networks based on GRU for sentence-level attention. In addition, similar studies can be conducted on different mental health issues.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ruba Skaik and Diana Inkpen designed and conducted the research; Ruba Skaik analyzed the data and implemented the classifiers; Ruba Skaik wrote the first draft of the paper; Diana Inkpen helped with the revisions of the paper; all authors had approved the final version.

Acknowledgment

This research is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] A. Abbe, C. Grouin, P. Zweigenbaum, and B. Falissard, "Text mining applications in psychiatry: A systematic literature review," *Int. J. Methods Psychiatr. Res.*, 2016.

[2] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on twitter," *Internet Interv.*, vol. 2, no. 2, pp. 183-188, May 2015.

[3] Z. Wang, G. Yu, and X. Tian, "Exploring behavior of people with suicidal ideation in a Chinese online suicidal community," *Int. J. Environ. Res. Public Health*, vol. 16, no. 1, 2019.

[4] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating machine learning algorithms for Twitter data against established measures of suicidality," *JMIR Ment. Heal.*, vol. 3, no. 2, p. e21, 2016.

[5] A. L. Nobles, J. J. Glenn, K. Kowsari, B. A. Teachman, and L. E. Barnes, "Identification of imminent suicide risk among young adults using text messages," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1-11.

[6] R. N. Grant, D. Kucher, A. M. León, J. F. Gemmell, D. S. Raicu, and S. J. Fodeh, "Automatic extraction of informal topics from online suicidal ideation," *BMC Bioinformatics*, vol. 19, no. 8, p. 211, 2018.

[7] A. E. Aladag, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu, and H. O. Bingol, "Detecting suicidal ideation on forums: Proof-of-Concept study," *J. Med. Internet Res.*, vol. 20, no. 6, p. e215, Jun. 2018.

[8] Z. Wang, G. Yu, X. Tian, J. Tang, and X. Yan, "A study of users with suicidal ideation on Sina Weibo," *Telemed. e-Health*, vol. 24, no. 9, pp. 702-709, 2018.

[9] M. Lv, A. Li, T. Liu, and T. Zhu, "Creating a Chinese suicide dictionary for identifying suicide risk on social media," *PeerJ*, vol. 3, p. e1455, 2015.

[10] Q. Cheng, T. M. Li, C. L. Kwok, T. Zhu, and P. S. Yip, "Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study," *J. Med. Internet Res.*, vol. 19, no. 7, pp. 1-10, 2017.

[11] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, and T. Zhu, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in *Proc. IEEE Int. Conf. Ubiquitous Intell. Comput.*, 2014, pp. 844-849.

[12] M. J. Vioulès, B. Moulahi, J. Azè, and S. Bringay, "Detection of suicide-related posts in Twitter data streams," *IBM J. Res. Dev.*, vol. 62, no. 1, pp. 7:1-7:12, 2018.

[13] A. Zirikly, P. Resnik, O. Uzuner, and K. Hollingshead, "CLPsych 2019 shared task: Predicting the degree of suicide risk in reddit posts," in *Proc. Sixth Work. Comput. Linguist. Clin. Psychol.*, 2019, pp. 24-33.

[14] E. Mohammadi, H. Amini, and L. Kosseim, "CLaC at CLPsych 2019: Fusion of neural features and predicted class probabilities for suicide risk assessment based on online posts," in *Proc. the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019, pp. 34-38.

[15] M. D. Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proc. CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2098-2110.

[16] A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," in *Proc. the First Workshop on Ethics in Natural Language Processing*, 2017, pp. 94-102.

[17] J. Jashinsky, *et al.*, "Tracking suicide risk factors through Twitter in the US," *Crisis*, vol. 35, no. 1, pp. 51-59, Jan. 2014.

[18] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. Conf. Weblogs Soc. Media*, 2014, pp. 216-225.

[19] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," in *Proc. 26th Int. Conf. Comput. Linguist. Tech. Pap.*, 2016, vol. 2, no. 1, pp. 3485-3495.

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[21] K. White, G. Li, and N. Japkowicz, "Sampling online social networks using coupling from the past," in *Proc. 12th IEEE Int. Conf. Data Min. Work.*, 2012, pp. 266-272.

[22] K. White, "Forecasting Canadian elections using Twitter," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, Cham., 2016, vol. 9673, pp. 186-191.

[23] R. Skaik and D. Inkpen, "Using Twitter social media for depression detection in the Canadian population," in *Proc. 3rd Artificial Intelligence and Cloud Computing Conference*, 2020.

Ruba Skaik is a PhD candidate at the University of Ottawa, in the School of Electrical Engineering and Computer Science. She has a M.Sc. and B.Sc. degree in Computer Science from Kuwait University. Her research interests include machine learning, natural language processing, and distributed database. She is an Oracle Certified Professional and has 15 years of experience in development, system analysis, and project management.

Diana Inkpen is a Professor at the University of Ottawa, in the School of Electrical Engineering and Computer Science. She obtained her Ph.D. from the University of Toronto, Department of Computer Science. She has a M.Sc. and B.Eng. degree in Computer Science and Engineering from the Technical University of Cluj-Napoca, Romania. Her research is in applications of Natural Language Processing and Text Mining. She organized seven international workshops and she was a program co-chair for the 25th Canadian Conference on Artificial Intelligence (AI 2012, Toronto, ON, May 2012) conference. She is the editor-in-chief of the Computational Intelligence journal and an associate editor for the Natural Language Engineering journal. She published a book on Natural Language Processing for Social Media (Morgan and Claypool Publishers, Synthesis Lectures on Human Language Technologies, the third edition appeared in 2020), 10 book chapters, more than 35 journal articles and more than 120 conference papers. She received many research grants, from which the majority include intensive industrial collaborations.