A Hybridized Feature Extraction Approach To Suicidal Ideation Detection From Social Media Post

Faisal Muhammad Shah
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
faisal.cse@aust.edu

Farsheed Haque
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
farsheed.haque@gmail.com

Ragib Un Nur
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
ragib08@gmail.com

Shaeekh Al Jahan
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
shaeekhshuvro@gmail.com

Zarar Mamud
Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
zarar26896@gmail.com

***Abstract*—Suicide's been a rising social problem. Concerned society has expressed worries regarding the recent increase in committing suicide. There are different stages of the suicidal act. If one can get recovery from early-stage which is, suicidal ideation, it is possible to reduce the number of suicides per year. Our study aims to detect suicidal ideation from social media-based using Natural Language Processing (NLP). We have applied the best feature extraction methods- Genetic and Linear Forward Selection (LFS) to select the best features from our feature vectors using our proposed formula. Moreover, We have created a robust feature set based on different computational and linguistic features. Finally, we have shown that by applying our hybrid feature extraction method we can get a significant increase in our accuracy to detect ideation.**

Keywords— Suicidal Ideation, Linguistic Feature, Wrapper Method, Filter Method

I. INTRODUCTION

Suicide is the act of terminating one's own life. The reasons behind committing suicide can be divided into two major parts- the first one is the impulsive decision and the other one is a mental disorder. For our research, the focus is on the lack of mental stability which leads to suicide. This starts with the thought of killing oneself from extensive thought and detailed planning. This is called Suicidal Ideation or Suicidal Thoughts. The main difference between suicidal ideation and suicidal behavior is that suicidal behavior indicates all possible acts which indicate activities that are intended to kill oneself [1]. The World Health Organization (WHO), says that in the whole world every year 800000 people commit suicide, which means one person dies because of suicide every 40 seconds [2].

We live in an age of modern era where people spend a good amount of time on social media which gives people a platform where they can anonymously express their thoughts. Again, this helps the people having suicidal ideation to share their sufferings. Also, online social groups like SuicideWatch [3] are created for this purpose where one can talk about their suicidal thoughts. We can use such online content to analyze the pattern and detect suicidal ideation from those content.
For our research purpose, we have collected textual posts from the subReddit called "Suicide watch". We have conducted a thorough analysis of the user content and we have taken both traditional features and linguistic features as linguistic features are always very important when we work with human sentiment and emotion. And, we proposed a hybrid method where the best feature selection methods are used to extract the best features from the feature set. Furthermore, we used our proposed method to create a robust feature set.

We show a complete procedure of suicidal ideation detection through machine learning using natural language processing known as NLP.

II. RELATED WORKS

There have been several works on suicidal ideation recently. Manas et al. [4] used the Columbia Suicide Severity Rating (CSSR) scale to perform four-level categorizations. They used Convolutional Neural Network (CNN), Feed Forward Neural Network (FFNN), Variations of Support Vector Machine (SVM), Random Forest to classify their gold standard dataset. Erin J. Lightman et al. [5] used linguistic features to extract suicidality from the text. They chose features like death-themed words, a reference to self and others, emotion words to differentiate suicidal lyrics from nonsuicidal. Their research finds that the average word count per song was larger for the non-suicidal songwriters. Shaoxiong Ji et al. [6] tried to predict suicidal ideation from online posts taken from Reddit and Twitter. They chose XGBoost, SVM, Random Forest, and Gradient Boosting Decision Tree (GBDT) as classifiers. They showed that adding new features helps to obtain more accuracy. Bilel Moulahi et al. [7] used a 4 class scale to annotate suicidal posts taken from twitter. They selected the twitter streaming API to collect posts. The classes to annotate are- no distress, minimal distress, moderate distress, and severe distress. For evaluating the approach different types of classification algorithms were used such as Naïve Bayes, Decision tree, K Nearest Neighbor Classifier (KNN).

III. DATASET

We have taken our data consisting of social media (Reddit) contents from Shaoxiong et all. [8].

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

Reddit is a social networking website. It contains groups based on a different area of interest which is known as subreddit. The people from different communities have built up different subreddits here. Our dataset user-posts containing suicidal ideation taken from subReddit called SuicideWatch. SuicideWatch is a community of Reddit where people express their suicidal thoughts, attempts, behavior, etc. The dataset consisted of 3549 numbers of data where a user-post column contains suicidal ideation. Also, there is another 3549 number of data of different popular Reddit posts that do not contain ideation. That means we have a total of 7098 data where the source is mentioned if each of them is from SuicideWatch or not. While making the dataset the researchers made sure that the users will remain anonymous. We have labeled our data according to the source. We have labeled '1' the data that are from SuicideWatch and '0' that are from other subreddits. After that, we have a properly annotated dataset ready to be classified by different supervised machine learning algorithms.

IV. METHODOLOGY

Fig. I show the workflow of our model. We have focused on computational features and Linguistic features and proposed our hybrid method to create a robust feature set and we used the feature to train the classifiers.

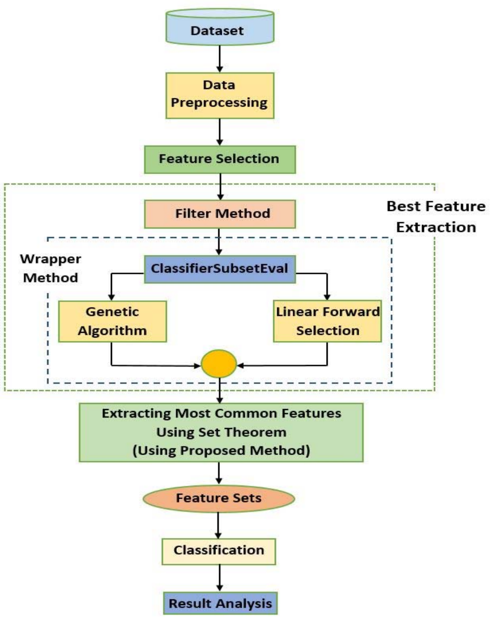

Fig. I: Model Diagram

A. Data Preprocessing

After collecting the dataset, each post was preprocessed before feature extraction. The posts were tokenized by words. For preprocessing we removed stop words and punctuations and we also applied stemming to reduce the word to its root.

B. Feature Selection

We have selected several features including TF-IDF, N-Gram (Unigram, Bigram, Trigram), and Other Linguistic Features Using LIWC 2015) [9].

a) TF-IDF: TF-IDF helps to reflect how important a term is in each document and it tells the weight of words in the given document. We used TF-IDF to measure the importance of words for both suicidal and non-suicidal user posts.

b) N-Gram: N-Gram is used to find the probability of n words in a given document. When the value of n=1 we get unigram, when it is n=2 we get bigram and when it is n=3 we get trigram. N-Gram is used in our model to predict suicidal word sequences for both suicidal and non-suicidal user posts.

c) LIWC: Linguistic Inquiry and Word Count (LIWC) program analyzes text and gives the statistics of the text on the linguistic level and gives results like the number of pronouns or emotion words etc that are used in a text [10]. To understand the suicidal thought process of a human being it is essential to analyze the user post on the human level and because LIWC analyzes a text using linguistic features it is essential to use the features extracted by LIWC on Suicidal Ideation detection problem. In Table I, we see the text analysis of an example done with LIWC.

An example

"I just want to end my life so badly. My life is completely empty and I don't want to have to create meaning in it. Creating meaning is pain. How long will I hold back the urge to run my car head first into the next person coming the opposite way. When will I stop feeling jealous of tragic characters like Gomer Pile for the swift end they were able to bring to their lives?"

TABLE I. LIWC FEATURE EXAMPLE

Label	All Pronoun	First Person Singular Pronoun	Positive Emotion	Negative Emotion	Present Tense Verb	Future Tense Verb	Death Related Words
1	13.33	9.99	2.67	6.67	12	4	1.34

C. Best Feature Extraction

Usually, a document contains hundreds or thousands of distinct words that are regarded as features that cause high dimensionality also many of the words can be noisy and redundant. This problem misleads the classifiers which eventually results in poor performance in general [11, 12]. To solve this problem we used the wrapper feature selection method. But the wrapper method has higher computational complexity so first, we used the filter method to get the best features to reduce the dimensionality of features. Then we used the wrapper feature selection method on the best features.

a) Filter Method: For a given work, the filter method of feature selection computes the usefulness of features. The method completes the selection by giving features with the highest score. Based on the work of Yang et al. [13] and Yarowsky et al. [14] we have selected information gain as the filter feature selection method for our model. Information gain is the amount of information that is gained by knowing the value of the attribute. It is the difference of entropy of the distribution before the break and the entropy of the distribution after the break. We have taken 150 best features for TF-IDF and N-Gram and 30 best features for LIWC according to information gain standard.

b) Wrapper Method: This feature selection method works out the informativeness of a feature by validating it with learning algorithms. By testing the combination of features it selects the optimum features for a particular task.

We have used linear forward selection and genetic algorithm (GA) based feature selection for our wrapper base as it is relatively faster for higher dimensions [15,16]. The genetic algorithm uses the concept of the biological evolution process of the survival of the fittest solution among many optimal solutions. Linear Forward Selection(LFS) wrapper method starts with no features in the features set and continues to add features one by one in the set of features until no improvement is detected in the model.

D. Hybridized Feature Extraction and Combination Method

We have proposed a unique method based on the set theorem to extract the most important features. Also, this method gives us many combinations for making feature sets.

$$[A2\,\{A1\,(F1)\}\,\cap\,A3\,\{A1\,(F1)\}] \cup [A2\,\{A1\,(F2)\} \cap A3\{A1(F2)\}] \cup \ldots \cup [A2\{A1\,(Fn)\}\,\cap\,A3\{A1(Fn)\}] \quad (1)$$

Here,
A1 = Information Gain
A2 = Genetic Algorithm
A3 = Linear Forward Selection
F = TF-IDF, Unigram, Bigram, Trigram, LIWC

In the proposed method (1) first, we ran Information Gain (A1) on each of our initially selected features (TF-IDF, Unigram, Bigram, Trigram, LIWC), this gave us 150 best features for TF-IDF, Unigram, Bigram, Trigram, and 30 best features for LIWC that has the highest score. Then we ran the Genetic Algorithm (A2) and Liner Forward Selection (A3) on each of the newly selected features and applied intersection to extract the common best features given by both of the algorithms. Finally, we applied union operation between the common best features to formulate a general method to create different combinations of feature sets. This proposed method is easy to implement and has less complexity.

E. Feature Set

Using the proposed method (1) we have created a feature set consisting of seven combinations. In Table II we have shown the different feature combinations with the number of features that we got after implementing our best feature extraction technique. The proposed method has selected the highest number of features that have the best impact on the outcome. Also for comparison, we have shown the number of features we got without implementing the best feature extraction technique. As, without the best feature extraction technique we get huge dimensions from normal features like unigram, bigram, etc. So in this case, we have used max feature selection and varied the max feature from 20-100 to reduce the dimensionality of the feature set. Through the trial and error process, we have found out that selecting 50 features gives us a good result in the classifier model.

TABLE II. FEATURE SET

Feature Sets	**Without Best Feature Extraction Technique**	**With Best Feature Extraction Technique**
TF-IDF + LIWC	50+94	33+17
Unigram + LIWC	50+94	36+17
Bigram + LIWC	50+94	19+17
Trigram + LIWC	50+94	13+17
Unigram + TF-IDF + LIWC	50+50+94	36+33+17
Bigram + TF-IDF + LIWC	50+50+94	19+33+17
Trigram + TF-IDF + LIWC	50+50+94	13+33+17

F. Classification

Suicidal Ideation detection is a binary supervised text classification problem. To make our model we introduced a binary-valued variable $Y_i = \{0, 1\}$ for every posts $P_i \in D$, dataset. $Y_i = 1$ denotes the post P_i shows suicidal ideation and $Y_i = 0$ means the post P_i does not show suicidal ideation. The job of the classification model is to determine whether any sentence in the post Pi has the keyword or the structure that indicates suicidal ideation. The model was trained by the feature sets to identify suicidal ideation in a given post. Traditional machine learning classifiers like Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest were used for classification.

V. RESULT ANALYSIS

We have used LIWC 2015 for linguistic analysis of the posts and taken linguistic features that are essential for suicidal ideation detection. For implementing Wrapper Method for extracting the best features we used WEKA version 3.8.4 software and finally for classification used Python 3.0 and scikit learn 0.22.2.

TABLE III. COMPARISON OF DIFFERENT CLASSIFIER USING DIFFERENT FEATURE SETS WITHOUT THE BEST FEATURE EXTRACTION

		Feature Sets						
Classifiers	Evaluation Metrics	TF-IDF + LIWC	Unigram+ LIWC	Bigram + LIWC	Trigram+ LIWC	Unigram + TF-IDF + LIWC	Bigram + TF-IDF + LIWC	Trigram + TF-IDF + LIWC
Naïve Bayes	Accuracy	0.528	0.532	0.634	0.532	0.529	0.539	0.532
	Precision	0.528	0.518	0.660	0.557	0.519	0.527	0.557
	Recall	0.753	**0.885**	0.551	0.316	0.756	0.765	0.316
	F1-Score	0.614	0.654	0.601	0.402	0.616	0.624	0.402
SVM	Accuracy	**0.700**	0.561	0.555	0.581	0.565	0.581	0.581
	Precision	**0.663**	0.567	0.567	0.584	0.569	0.588	0.584
	Recall	0.811	0.515	0.464	0.561	0.540	0.537	0.561
	F1-Score	**0.730**	0.539	0.510	0.572	0.554	0.561	0.572
KNN	Accuracy	0.669	0.662	0.652	0.663	0.668	0.665	0.663
	Precision	0.625	0.632	0.620	0.617	0.639	0.634	0.617
	Recall	0.848	0.696	0.786	0.858	0.774	0.781	0.858
	F1-Score	0.719	0.696	0.693	0.718	0.700	0.700	0.718

TABLE III. (Continued).

Random Forest	Accuracy	0.642	0.634	0.622	0.616	0.637	0.612	0.616
	Precision	0.627	0.619	0.615	0.606	0.621	0.610	0.606
	Recall	0.700	0.697	0.658	0.660	0.701	0.625	0.660
	F1-Score	0.661	0.634	0.635	0.632	0.658	0.617	0.632

TABLE IV. COMPARISON OF DIFFERENT CLASSIFIER USING DIFFERENT FEATURE SETS WITH THE BEST FEATURE EXTRACTION

		Feature Sets						
Classifiers	Evaluation Metrics	TF-IDF + LIWC	Unigram+ LIWC	Bigram + LIWC	Trigram+ LIWC	Unigram + TF-IDF + LIWC	Bigram + TF-IDF + LIWC	Trigram + TF-IDF + LIWC
Naïve Bayes	Accuracy	**0.736**	0.657	0.717	0.554	**0.736**	**0.736**	0.640
	Precision	0.688	0.687	0.660	**0.705**	0.688	0.688	0.696
	Recall	0.866	0.576	0.897	0.181	0.866	0.866	0.498
	F1-Score	**0.767**	0.626	0.761	0.284	**0.767**	**0.767**	0.580
SVM	Accuracy	0.716	0.717	0.717	0.716	0.716	0.716	0.716
	Precision	0.662	0.662	0.662	0.662	0.662	0.662	0.661
	Recall	0.885	0.885	0.885	0.885	0.885	0.885	0.884
	F1-Score	0.757	0.757	0.758	0.757	0.757	0.757	0.757
KNN	Accuracy	0.712	0.713	0.712	0.712	0.712	0.712	0.712
	Precision	0.660	0.660	0.660	0.660	0.660	0.660	0.660
	Recall	**0.877**	**0.877**	**0.877**	**0.877**	**0.877**	**0.877**	**0.877**
	F1-Score	0.753	0.753	0.754	0.753	0.753	0.753	0.753
Random Forest	Accuracy	0.729	0.729	0.726	0.727	0.729	0.729	0.730
	Precision	0.679	0.678	0.674	0.676	0.679	0.679	0.679
	Recall	0.871	0.873	0.873	0.870	0.871	0.871	0.874
	F1-Score	0.763	0.763	0.761	0.761	0.763	0.763	0.764

This section compared various classification methods using different features sets. The classifiers are Naïve Bayes, SVM (Support Vector Machine), K- Nearest Neighbor, and Random Forest. And we used Accuracy, Precision, Recall, and F1-Score as evaluation metrics. Table III shows the comparison of classifiers without the best feature extraction, here we got the highest accuracy which is 0.700 in KNN. No other classification method reached seventy percent which is not satisfactory at all. This happened due to the high dimensionality of the feature vector.

In Table IV we have shown the comparison of classifiers with our proposed best feature extraction technique, we got better accuracy above 0.7 for most of the classifiers and achieved 0.736 for Naïve Bayes classifier. As in sentiment analysis, every word is taken as a feature, and people tend to have words to express themselves that make a huge feature. For this Naïve Bayes performs well as it takes independent feature assumption. The above comparison in Table III and Table IV indicates that the best feature extraction technique surely works better in suicidal ideation detection than taking max features in machine learning models.

VI. CONCLUSION

The reduction of the rate of suicide has become a crying need for our society. We hope, our contribution will help the community to detect suicidal ideation at an early stage. We have analyzed text using traditional machine learning to detect suicidal ideation. Later, we showed that extracting the best features reduces complexity as well as increases accuracy. In the future, Deep learning techniques can play a vital role to improve accuracy by selecting features based on a better understanding of the sequential relationship of sentences.

REFERENCES

[1] Ramit Sawhney, Prachi Manchanda, Raj Singh, and Swati Aggarwal "A Computational Approach to Feature Extraction for Identification of Suicidal Ideationin Tweets" Proceedings of ACL 2018, Student Research Workshop, pages 91–98 Melbourne, Australia, July 15 - 20, 2018

[2] World Health Organization. Mental Health. Suicide data. Retrieved February 16, 2020, from https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/

[3] SuicideWatch, "https://www.reddit.com/r/SuicideWatch/" [Accessed: 26-Feb-2020]

[4] Manas Gaur et all. "Knowledge-aware Assessment of Severity of Suicide Risk for Early Intervention" in *IW3C2 (International World Wide Web Conference Committee*), San Francisco, USA, 2019

[5] Erin J. Lightman, Philip M. McCarthy, David F. Dufty, Danielle S. McNamara "Using Computational Text Analysis Tools to Compare the Lyrics of Suicidal and Non-Suicidal Songwriters", 2007

[6] Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long "Supervised Learning for Suicidal Ideation Detection in Online User Content" Hindawi Complexity, 2018

[7] Bilel Moulahi, Jérôme Azé, Sandra Bringay Detection of Suicide-Related Posts in Twitter Data Streams

[8] Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long "Supervised Learning for Suicidal Ideation Detection in Online User Content" Hindawi Complexity, 2018

[9] LIWC, "https://liwc.wpengine.com/" [Accessed: 26-Feb-2020]

[10] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates, 71(2001):2001. *Short-term group psychotherapy with suicidal patients.* (2nd ed.) Durkheim, E. Suicide. Farberow, N.L., & Marmor, 1951

[11] Liu H, Yu L "Toward integrating feature selection algorithms for classification and clustering" IEEE Trans Knowl Data Eng 17(4):491–502, 2005

[12] Sebastiani F "Machine learning in automated text categorization" ACM Comput Surv 34(1):1–47, 2002

[13] Y. Yang, J.O. Pedersen, "A comparative study on feature selection in text categorization" Proceedings of the 14th International Conference on Machine Learning (ICML), pp. 412–420, 1997

[14] D. Yarowsky, G. Ngai, R. Wicentowski "Inducing multilingual text analysis tools via robust projection across aligned corpora" Proceedings of the First International Conference on Human Language Technology Research, pp. 1–8, 2001

[15] Xuelian Deng, Yuqing Li, Jian Weng " Feature selection for text classification: A review" Multimed Tools Appl **78,** 3797–3816, 2019

[16] Serkan Gunal "Hybrid feature selection for text classification" Turkish Journal of Electrical Engineering and Computer Sciences, 20, 1296-1311, 10.3906/elk-1101-1064, 2012