IJCI V10-2 (2023) 1–15
International Journal of Computers and Information (IJCI)
Available online at *https://ijci.journals.ekb.eg/*

A New Approach to Suicide Ideation Detection from Text Content

Abdallah Basyouni, Hatem Abdelkader, Asmaa Aly
Information Systems dept., Faculty of Computers and Information, Menofia University, Egypt.
abdallah.basyounia@ci.menofia.edu.eg, hatem.abdelkader@ci.menofia.edu.eg, asmaa.elsayed@ci.menofia.edu.eg

Abstract

Suicide is a serious issue in modern society all over the world. Suicide can be caused by several risk factors. Anxiety, hopelessness, social isolation, and depression are the most popular risk factors. Early detection of those risk factors can help reduce or prevent the number of suicide attempts. Many expressions of suicidal thoughts can be discovered in online communities, mostly by young people. In this paper, a new approach to detecting suicidal ideation is built using natural language processing (NLP), and machine learning techniques. This study compares three classifiers, Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB). Our study extracted various feature sets, namely, Statistical, TFIDF, POS, N-grams, and Topic Modeling features. Various feature reduction techniques were used, Principal Component Analysis (PCA) and Information Gain (IG). The study aims to increase the suicide ideation detection accuracy, and address shortcomings in previous studies such as using few feature sets, focusing only on the level of words without considering the meaning context, and using all extracted features in the classification task, which includes some irrelative and redundant features. In this study, the RF classifier achieves the highest classification accuracy of 97.02% when using PCA as a future reduction technique. This study proved that using expressive feature sets and selecting relevant and informative features can achieve a more accurate classification process.

Keywords: Suicide Ideation; Reddit Social Media; Machine Learning; Feature Extraction; Feature Selection;

1. Introduction

Mental illness is the fifth most common illness in the world [1]. The economic cost of treating mental disorders was estimated to be $2.5 trillion in 2010 and is expected to double by 2030 [2]. One of the primary objectives of the World Health Organization's (WHO) Comprehensive Mental Health Action Plan 2013–20 was to create robust information systems for mental health, involving boosting the ability for people's health diagnoses [3]. According to the data point of the WHO, suicide causes more than 800,000 deaths each year, every 40 seconds there is a suicide case committed, and nearly 30% of those who committed suicide note their thoughts previously [4]. According to various reports, the prevalence of suicidal ideation among the world population is approximately 9%, and this is higher for people aged 18-25 years [5]. At-danger people may be identified as suicide planners and suicide attempters, the rapport between these two classes is commonly a topic of debate in analysis and research communities [6]. The most frequently cited risk factors (frustration, depression, hopelessness, anxiety) associated with suicide are predictors of suicidal thoughts [7]. Suicidal thoughts are people's intentions to end their lives. Detecting suicidal thoughts early is one of the best strategies to save people's life. Suicidal people may leave their thoughts and suicide plans in the form of suicidal notes. Detecting suicidal thoughts is discovering those risky thoughts before tragedy strikes.

There is a rapid growth of social media or social networks such as Twitter, Weibo, Reddit, and Facebook. With this growth, there is an increasing desire among people to create online communities and connect to share their thoughts. Psychiatric patients do not prefer going to a psychiatrist because they wish to remain anonymous. With the possibility of the user hiding his identity on social networks. It made him comfortably express his suicidal thoughts without fear of being found out by anyone [8–10]. Online checking tools work as standard appreciation strategies, such as health monitoring tools that can analyze indicators of mental disorders [11]. Prior research on suicide analysis and prohibition mostly focuses on its psychological and clinical sides [12].

Lately, several studies have headed to natural language processing approaches and machine learning [13]. These studies have utilized the "International Personal Examination Screening Questionnaire," and analyzed suicide posts from social media. However, these studies have their limitations. First, from the psychological perspective, gathering data on patients is very expensive, some online data may assist in understanding suicidal thoughts. Second, plain feature sets are not sufficient to detect suicidal ideation, leading to the development of new ways of detecting it. There is a substantial body of work addressing mental health issues through social media content. TeenLine, Tumblr, Instagram, Twitter, and Reddit have all been used as data sources for computational social science research [14–16]. Social media have become a big repository for users' thoughts. Social media content enables us to analyze people's opinions and thoughts.

Reddit is an online discussion forum where people express their opinions, feelings, thoughts, and family problems. Some subreddits permit people to talk about a specific topic. "SuicideWatch (SW)" is one of the subreddits where people can talk and share their potential suicide thoughts and intentions [17].

Feature selection is selecting a subset from the main features by banishing redundant, irrelevant, or noisy features. Feature selection usually leads to higher learning performance. According to different search strategies, feature selection is classified into three methods. (1) Wrapper methods, (2) filter methods, and (3) embedded methods [18]. Filter methods don't use classification algorithms to remove irrelevant features. They weigh for features by analyzing the main characteristics of the data according to statistical standards of the used filter strategy [19]. PCA minimizes the universal features set without missing much information by maintaining most of the main variability of data [20]. Information gain calculates the importance of a given feature by computing its entropy value (how much it is informative) about the class [19]. Redundant and irrelevant features mislead the learning algorithm, raise the time complexity, and drain computation resources [21].

The main purpose of this paper, introducing solutions to the early detection of suicidal ideations in Reddit users' posts through natural language processing techniques and effective machine learning approaches. With a comprehensive analysis of the posts, the language patterns, and the topic modelling to understand suicidal ideation from the perspective of textual data analysis. Five different robust sets of effective features are extracted from text data with n-gram analysis. In the sectors of probability and computational linguistics, an n-gram (also known as Q-gram) is a contiguous sequence of n items from a given sample of text or speech. The items can be letters, phonemes, syllables, words, or base pairs according to the application. The n-grams generally are collected from a text or speech corpus. N-grams may also be called shingles items are words [20]. Two filter feature selection methods were developed for the original feature set. Three supervised learning algorithms were compared to detect suicidal ideation on Reddit social media. It is a robust model of automatic suicidal ideation detection on Reddit social media content with the selection of relative or informative features from the original features set.

Our study has a specific set of contributions.

- **N-gram analysis**: As an informative feature set, n-gram in the form of a trigram was used. N-gram represents a contiguous sequence of three words. The trigram was used to keep the context and intended meaning of the users' posts. For example, suppose two documents (two posts) contain "I want to kill myself" and "I don't want to kill myself". If unigram or BOW are used, only the existence of each word individually is discussed. In this case, there is no difference between the two expressions even though they are opposite in meaning. Unlike the n-gram which preserves context and considers the presence of words in a specific window. In our case, the window is represented by three words which are called trigrams.

- **Feature selection**: The previous studies extract a set of features. They use all features as input to machine learning algorithms without thinking that some of these features may be irrelevant and misleading for the algorithm. In our study, two filter selection methods were applied. Principal component analysis (PCA) and Information gain (IG) select only the relevant features from the original features set as input to the classification algorithms. Using the selected features reduces the testing time and increases the accuracy of detecting suicidal ideation. The time of detecting the suicidal person is a very important factor in preserving a suicidal user's life before occurring tragedy.

This paper is structured as follows: Section 2 introduces the background and related works on the detection of suicide ideation. Section 3 examines our proposed methodology. For new features set and feature selection task with its impact. Section 4 discusses the experimental results as well as the most robust machine-learning approaches for detecting suicidal ideation. Section 5 concludes this paper and presents future work directions.

2. Background Knowledge and Related Works

2.1 Background Knowledge.

Because of the increase in the number of suicides in the past years, researchers have tended to reveal suicidal intentions on social media. Many reasons for suicide are classified as complex, and they are a complex interaction of several factors [22]. The research techniques used to detect suicide also include many areas and approaches. For example, clinical methods may check the resting heart rate [23] and event-linked instigators [24]. Traditional approaches also include the use of questionnaires to assess risks and potential threats to suicide and class interactions with the patient [25]. The purpose of text-based suicide classification is to determine whether users, through their online content, have suicidal thoughts. These methods include filtering out words related to suicide [26], [27] and phrase filtering [28]. Machine learning methods especially supervised learning and natural language processing have also been used in this area. Context features, n-gram features, class-specific, syntactic, and knowledge-based features [29].

2.2 Related Works

Recently, more studies have been applied to data collected from social media. Cash et al. [30] and Shepherd et al. [31] have conducted psychology-based data analysis for content that mentions suicidal ideation on Twitter and Myspace social media. But these studies suffer from the problem of gathering data and/or patients are very complex and expensive. Reddit has caught the attention of researchers, too. Bashir and Huang used linguistic indicators to examine the reply bias [32]. Choudhury et al. [33] implemented a statistical approach based totally on a score-matching model to elicit some of the hallmarks that reveal the transition from mental health discourse to suicidal thoughts. This transition can be through three phases: thinking, contradiction and decision-making. The first phase contains the risk factors for suicide such as hopelessness, anxiety, and distress. The second phase is linked to self-loathing or self-flagellation and social isolation. The third phase is related to aggressiveness and committing suicide plan. But this study has some limitations because Reddit enables the use of semi-anonymous identities, including having multiple accounts, and a shift from MH to SW communities may occur via such an account. Such shifts would not be captured by the data collection process. Users with a history of mental illness may also directly post on SW without ever posting on any mental health subreddits. Colombo et al. [34] studied suicidal ideation in suicidal users' tweets based on their behavior in social media interactions which led to a high level of strong connectivity between users and analyzing this connectivity among suicidal users. But this study suffers from some limitations, their analysis was conducted on a limited size set of annotated posts, furthermore, the tweets classified as inclosing suicidal thoughts did not appear to be included in large percentages (only about 10% of tweets collected using suicide-related keywords). Kumar et al. [35] studied the impact of celebrity suicide and their suicide posts on social media platforms. And users imitate them, which is what is called the Werther effect. But this experiment suffers from some limitations, first, being limited to those people who participate in the SuicideWatch, there may be a self-selection bias in this population. Second, it is also a group who are selecting an online platform for asking for help, instead of other (offline) modalities of suicide support.

Recognition of regular language styles in the textual content of social media platforms enables more accurate discovery of suicidal ideation, by applying several machine learning approaches to different NLP techniques. Braithwaite et al. [36] showed that machine-learning techniques are more effective in distinguishing suicidal people from others. But this study suffers from, only limited amounts of clinically significant suicide, not exceeding 200 individuals, that were analyzed. Sueki et al. [37] examined the suicidal thoughts of Twitter users in their twenties in Japan and mentioned that language patterns and preferences are important for detecting

suicidal indicators in textual data. For example, the phrase "want to suicide" is more often related to suicidal ideation than the phrase "want to die". This study suffers from some shortcomings, including the selected participants being young Japanese people; the results may vary for youths of other cultures. O'Dea et al. [38] determined the level of concern for Twitter posts related to suicide using SVM and logistic regression on TFIDF features, this study suffers from some shortcomings, the analyses used to extract this model were primarily based on single words, and plain feature set was used (only TFIDF). Okhapkina et al. [39] studied the tuning of information retrieval strategies to determine the harmful informational effect on social media platforms. He created a lexicon of terms related to suicidal content. He applied singular vector decompositions and TF-IDF metrics for them. Ji et al. [40] presented a new dataset for suicide ideation and an approach for early detection of suicide ideation on Reddit and Twitter social media, but this approach suffered from focusing only on the level of words without the meaning context, as well as using all the extracted features in the classification task, which contains some irrelative and redundant features. These features will mislead the learning algorithm and increase computational complexity and time. Michael et al. [41] developed a machine learning-based classification approach to detect suicidal thoughts on Reddit social media, which suffers from some limitations, as few feature sets were used, it was not reasonably representative of the data and focused only on the word level rather than the meaning. Renjith et al. [42] propose a combined LSTM-Attention-CNN model for analyzing social media submissions to detect any underlying suicidal intentions, this model demonstrated an accuracy of 90.3%. This study suffers from ignoring many meanings and context-related features. Liu et al.[43] Suggesting an ensemble method based on feature fusion to detect suicide ideation on Weibo social media, this study extracts a set of features, namely basic statistical characteristics (BSC), risk factors for suicide (RFC), and word embedding clustering (WEC). This study has some limitations as well. Due to Weibo's privacy settings, users' age, gender, location, and other information cannot be obtained. Users in Weibo communities are primarily young people. As a result, the data used has some bias, and the occurrence of suicidal behavior as well as the relationship between age and gender have been reported.

Feature selection is a procedure that includes excluding irrelevant and duplicated features from a feature set to enhance the accuracy of machine learning algorithms. Sameen et al. [44] develop a methodology to improve the text classification accuracy using feature selection, using Random forests-based feature selection (RFFS), and better accuracy than the original feature set was obtained. Kumar et al. [45] studied kidney Ultrasound Images Using GLCM and PCA and the results proved that GLCM in combination with PCA for feature reduction achieves high classification accuracy when classifying images using Artificial Neural Networks (ANN). Md. Palash Uddin et al. [46] classified Hyperspectral Remote Sensing Images based on PCA feature selection and other methods, and the results showed that the classification performance using the feature selection is better than the classification performance using the original features set. Erick et al. [47] implemented a hybrid filter model for feature selection based on information gain and principal component analysis, the results illustrated that the hybrid filter model selects relevant feature sets, reduces data dimensions, and decreases training and testing time, thus providing better classification performance that measured by accuracy.

3. Proposed Algorithm

This study aims to improve the detection of suicidal ideation on social media. This study proposes a novel approach to detecting suicidal ideation. It depicts a model with a series of successive phases. Data collection, data pre-processing/cleaning, feature extraction, feature selection, classification process, and results analysis are the steps involved. Fig 1 depicts an overview of our proposed model, which includes data pre-processing and the extraction of a robust set of features using NLP techniques such as (Statistical, TF-IDF, POS, Tri-gram and Topic Modeling Features). Then, using PCA and Information Gain (IG), the features selection method was used to pass only relevant or informative features. Finally, using the selected feature set, apply machine learning algorithms to detect suicidal thoughts based on input features and analyse the results.

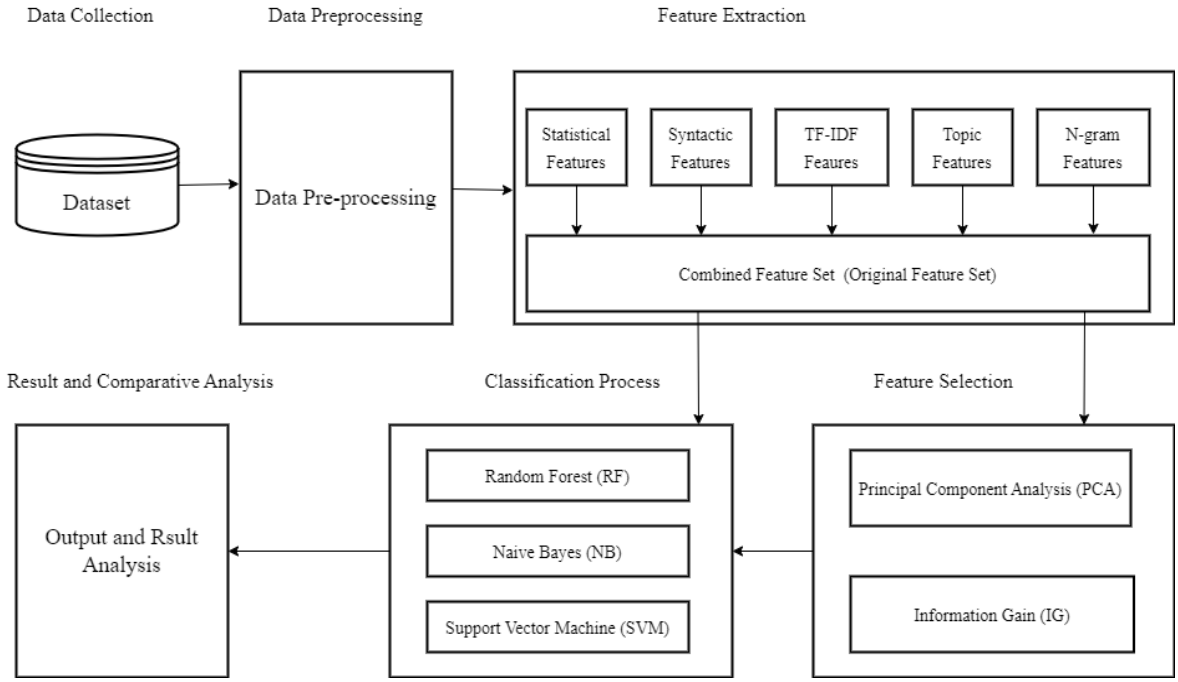


Fig 1: Proposed Model

3.1. Data Collection

3.1.1 Dataset

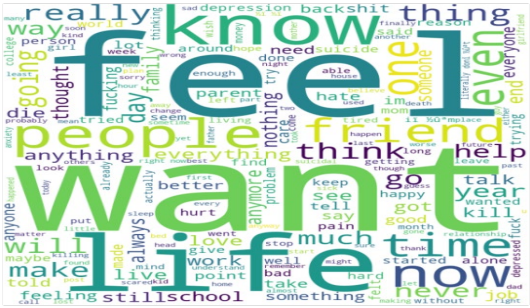
For training our classification models, a dataset collected from Reddit social media was used, where users can write their opinions, feelings, and thoughts in the form of links, text posts, or voting posts. Ji et al. [40] compiled the dataset used in this study, including a list of suicide and non-suicidal posts. A unique ID is used instead of their personal information to protect the users' privacy. Because users tend to participate in various types of subreddits, each group contains a similar random number of messages drawn from various topics. Our dataset was compiled from 3652 non-suicidal posts and 3549 suicide-reporting publications from relatively large subreddits dedicated to assisting at-risk individuals. Suicidal posts were gathered from the "SuicideWatch" subreddit (https://www.reddit.com/r/SuicideWatch/). Other popular subreddits (https://www.reddit.com/r/all/, https://www.reddit.com/r/popular/) provided posts with no suicidal intent. Non-suicidal data collection is entirely user-generated content, with no news aggregation posts or administrators. Non-suicidal posts are generated from subreddits that are thematically relevant to family and friends. Table 1 displays examples from suicide and non-suicide posts.

Table 1: Examples of Suicide posts and Non-Suicide Posts on Reddit social media

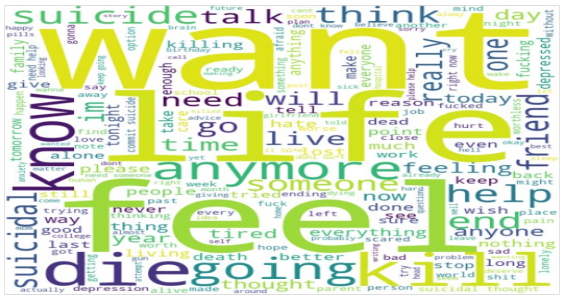
Suicide Posts	**Non-Suicide Posts**
I think I'm close to my end. can someone just convince me not to commit suicide?	I covered all my weapons in glue. But I've seen a ton of bars and restaurants demanding his freedom lately.
I want to die. I've wanted to die for years. I have nothing to give. I just have no will to live anymore I've lost everything.	There are eleven types of people in this world... He was fed up with other people.
I have no life or friends to celebrate with. I would just accept not being so suicidal and depressed all the time as a gift.	Your vacuum cleaner has been gathering dirt on you for years. I don't like political jokes, too many of them get elected.

3.1.2 Word Cloud

word clouds that are used to present a visual representation to understand the data. The title and text of Reddit users' posts indicating a potential suicide risk are shown separately in Fig 2(a) and 2(b). As shown, suicide posts often use words like "die", "life", "kill", and "suicide", which provide a direct indication of users contemplating suicide. Words expressing intentions or feelings are used frequently such as "want", "know" and " feel ". For example, suicidal posts involve sentences like "I don't want to live anymore I hate this world", "I want to die" and "I want to end my life" as indicated in Table 1.



(a) Title Word Cloud



(b) Body Word Cloud

Fig 2: Word Cloud Visualization of Reddit text posts

3.2. Data Pre-processing

Data pre-processing is a major step in all machine learning tasks. The data collection is mostly an uncontrolled process, leading to the production of data suffering from shortcomings and problems such as missing values, out-of-range values, noised data, etc. Deciding based on poorly examined data produces misleading results. Using pre-processed data can improve the model's performance. Pre-processing consists of a series of filtering phases on Reddit posts to convert row data into an appropriate format understood by machine learning models. In our experiment, the Natural language toolkit (NLTK) and Spacy were used to pre-process our dataset before it moved into the training stage. The beginning with the case conversion step where all words were converted to lowercase. Next, The duplicated posts were removed from the dataset. Then, tokenization is used to split the user's posts into individual words or tokens. The HTML tags were stripped from the data. This step followed isby removing accented characters for example ''Sómě Áccěntěd těxt'' converted into ''Some Accented text''. Then, expanding contraction step was applied, in this step, first, a contraction map containing all the possible contractions and possibly expansion were formulated, for example ("haven't": "have not", "he'd": "he had / he would", "he'd've": "he would have", etc.), then all the tokens in each post were compared and transformed with the contraction map into their expended contraction if possible. Then the special characters were removed (any character except a-z, A-Z and 0-9). Next, the repeated characters were removed. Custom stop words were removed, where NLTK and Spacy stop words were listed, but some stop words which could change the meaning of the sentence were excluded like ''not'', ''myself'', ''alone'', ''against'', ''nobody'', ''still'', etc. This step was followed by removing one-letter words, Then, Name Entity Recognition (NER) was removed. Applying the lemmatization where can't drop word ends could produce meaningless word parts like stemming. Finally, the cleaned data is ready for the feature extraction phase.

3.3. Feature Extraction

Feature extraction is the process of converting the cleaned data into an understandable format for learning algorithms. Five robust feature sets were extracted, namely TF-IDF, POS, trigram, topic modeling and statistical features. The implementation details of the feature extraction phase are described below in pseudo-code representation, as shown in Algorithm 1. Where five sets of features were extracted from each post and combined them into a single set known as the original set. Some notations are provided below to aid in the legibility of the proposed algorithm.

1. **DOCs:** {doc1, doc2, …, docn}: the posts of users in our dataset.
2. **Fet:** variable that stores the temporary result.
3. **STATISTICAL_SET**: a list that contains statistical features.
4. **POS_SET**: a list that contains syntactic features.
5. **TF-IDF_SET**: a set that contains tf-idf features.
6. **TOPIC_SET**: a set that contains topic modelling features.
7. **N-GRAM_SET**: a set that contains n-gram features.
8. **ORIGINAL_SET**: a set that combines all extracted features.
9. **getStatistics**: a function to generate statistical features.
10. **getPos**: a function to generate pos features.
11. **getTopics**: a function to generate topic features.
12. **getTFIDF**: a function to generate tf-idf features.
13. **getNgram:** a function to generate n-gram features.
14. **merge:** a function that combines several lists into a single set.

Algorithm 1: Feature Extraction phase Algorithm

```
Input: DOCs
Output: ORIGINAL_SET
1  begin
2     foreach DOC ∈ DOCs do
3        Fet ← getStatistics (DOC)
4        STATISTICAL_ SET.push(Fet)
5        Fet ← getPos (DOC)
6        POS_ SET. push (Fet)
7        Fet ← getTopics (DOC)
8        TOPIC_ SET. push (Fet)
9        Fet ← getTFIDF (DOC)
10       TF-IDF_ SET. push (Fet)
11       Fet ← getNgram (DOC)
12       N-GRAM_ SET. push (Fet)
13    end foreach
14    ORIGINAL_ SET ← merge (STATISTICAL_ SET, POS_ SET,
                      TOPIC_ SET, TF-IDF_ SET, N-GRAM_ SET)
15    return ORIGINAL_ SET.
16 end
```

3.3.1 *Complexity Analysis.*

To evaluate the complexity of algorithm 1, the CPU time required to execute each statement is computed in the term of Big O, as shown in Table 2. Where n is the number of documents (posts) in our dataset and k is the number of tokens in each document. All constants are neglected when calculating the Big O. The total Big O in algorithm 1 is the product of loop Big O multiplied by the sum of Big O of all statements inside the loop.

Table 2: Complexity Analysis of Algorithm 1

Statement	Time Complexity (Big O)
Loop	*O (n)*
push	*O (1)*
getStatistics	*O (1)*
getPos	*O (k)*
getTopics	*O (n)*
getTFIDF	*O (1)*
getNgram	*O (k)*
Merge	*O (1)*
Total running time	*O (**n^2** + kn)*

3.3.2 *Statistical Features.* user posts on Reddit social media are different in length where some posts use short phrases, but others use long and complex phrases, and some statistical feature sets can be extracted from their posts. First, user posts were tokenized, statistical features were extracted as follows.

- For title: number of tokens, words, and characters.
- For body: number of tokens, words, characters, sentences, and paragraphs.

3.3.3 *POS Features.* refer to the part of speech features also called syntactic features. This set of features is useful in natural language processing (NLP) tasks. This set of features was extracted as a part of our features sets to present similar syntactic properties in the users' posts on Reddit social media. Popular POS tags involve verbs, nouns, adverbs, participles, pronouns, articles, and conjunctions. Each post was parsed and tagged, counting the number of each category in both text and title.

3.3.4 TF-IDF Features. Term Frequency Features. TF-IDF is a statistical measure that assesses how closely a word is related to a document in a set of documents. This set of features was extracted to measure the weight of different words from both non-suicidal posts and suicidal posts, this is done by multiplying two metrics: how many times a word occurs in a document, and the inverse document frequency of the word in a group of documents. The TF-IDF features were used to assess the importance of words in suicidal and non-suicidal user posts. The equation below explains TF-IDF.

The term frequency (tf) of the term t in document d is computed as follows.

$$tf(t,d) = \frac{C_{t,d}}{\sum_k C_{t,d}} \quad (1)$$

Where $C_{t,d}$ is the number of occurrences of t in d and $\sum_k C_{t,d}$ is the total number of terms in d. The inverse document frequency (idf) is also calculated as follows.

$$idf\ (t,D) = \log\frac{D}{d_t} + 1 \quad (2)$$

Where D is the total number of posts in the dataset and d_t is the number of posts in the dataset that contains t. Finally, the tf-idf is computed by multiplying Eq. 1 by Eq. 2.

$$tf - idf\ (t,d,D) = tf(t,d) \times idf(t,D) \quad (3)$$

3.3.5 *Topic Features:* Both non-suicide posts and suicide posts that talk about different topics can provide a good understanding of the two categories. In our experience, a latent Dirichlet task (LDA) [38] was applied to detect ambiguous topics or topics in user-generated posts. Each topic is a combination of the probability of a word appearing in the topic, and each user post is a mixture of the probability of the topics. In our experiment 10 topics extracted

3.3.6 *N-gram Features*. N-gram was used as one of our work contributions. N-gram was used in the form of n-gram to preserve context and the indented meaning from the user's post on Reddit social media, by studying the presence of words next to each other in the user's post instead of studying the presence of the word alone. N-grams of texts are widely used in natural language processing and text mining tasks. They are originally a group of co-occurring words or tokens within a given window (in our experiment the window equals three words) and when computing the n-grams you usually move one word forward. For example, for the sentence "I

want to kill myself.”. If N or window size =3 (known as tri-grams, then the n-grams would be: (i want to, want to kill, to kill myself).

One of the biggest shortcomings of previous studies in detecting suicidal thoughts is that they do not care about the intended meaning of the sentence. Previous work only examined the presence of single words in a user's post, but what about having the words next to each other? As well as what about the order of the sequential words? Previous works assume that the existence of the sentence "I want to kill myself" is the same as "I do not want to kill myself" because it does not concern itself with the words next to each word or the order of the words.

3.4. Feature Selection

Once the feature extraction process was completed, the number of features extracted was large. A large number and variety of features lead to misleading the learning algorithm, which leads to decreasing the classification performance due to the existence of many features that have no value (irrelevant) or are duplicated. To resolve this issue, two feature reduction techniques were applied for selecting the most informative and relevant subset of features based on specific measurements. The main objective of feature reduction is to increase classification performance, as well as increase processing speed and memory usage A brief overview of the two used feature reduction techniques. The implementation details of the feature selection phase are described below in pseudo-code representation, as shown in Algorithm 2. where the original set (output of the feature extraction phase) was used as input to PCA and IG. Two feature sets are produced because of the feature selection phase: the principal components by PCA and the selected features by IG. now there are three feature sets (Original set, principal components, and IG features) to feed into the classification phase. Some notations are provided below to aid in the legibility of the proposed algorithm.

1. **ORIGINAL_SET**: a set that combines all extracted features.
2. **PCA_LIST**: the principal components generated by PCA for ORIGINAL_ SET.
3. **IG_LIST**: the chosen features by IG for ORIGINAL_ SET.
4. **getPrincipalCompinents:** a function that computes the principal components using the PCA algorithm.
5. **getBestFeatures:** a function that selects the best features using IG.

Algorithm 2: Feature Selection phase Algorithm

Input: *ORIGINAL_SET*
Output: *PCA_SET and IG_SET.*

```
Begin
    PCA_SET ←getPrincipalComponents (ORIGINAL_SET).
    IG_ SET ←getBestFeatures (ORIGINAL_SET).
    return PCA_SET, IG_SET.
end
```

3.4.1 Complexity Analysis.

To evaluate the complexity of algorithm 1, the CPU time required to execute each statement is computed in the term of Big O, as shown in Table 3. Where *n* is the number of documents (posts) in our dataset and *f* is the number of features that represent each post. All constants are neglected when calculating the Big O.

Table 3: Complexity Analysis of Algorithm 2

Statement	Time Complexity (Big O)
getPrincipalComponents	$O(n^3)$
getBestFeatures	$O(f^2)$
Total running time	$O(n^3 + f^2)$

3.4.2 Principal Component Analysis (PCA). The principal component analysis is applied for dimensionality reduction when many features are used, and the components are highly correlated. PCA produces a smaller set

of synthetic variables which will act as the variance of a set of the observed variable. The principal components are the calculated synthetic variables. The variables are perpendicular to PCA and the principal components with the biggest variation are removed from the dataset. The PCA applied as follows on the dataset.

- a dataset that has a mean of zero is made by subtracting the mean of the data from each data dimension, which is also called standardizing the range of continuous initial variables.
- Calculate a covariance matrix for identifying correlations.
- Calculate Eigenvectors and Eigenvalues of the covariance matrix for identifying principal components.
- The highest Eigenvalues represent the principal components, and the Eigenvalues of less importance are eliminated and create a feature vector.
- Paraphrasing data based on principal components axes.

Instead of using the original feature set of 200 features as input to the classification models, by using PCA, only 67 components were produced.

3.4.3 Information Gain (IG): The IG is a common measurement that for defining the scope to which a particular feature produces informative value about a class. It gives a ranked score for each feature, then chooses the highest score and neglects the lower scores. Instead of using the original feature set of 200 features as input to the classification models, IG was applied to it with a threshold value and get only 91 features. Table 4 displays the original features as well as the features chosen by IG. There were fewer features used, but they were the most important ones.

Table 4: Original Features Set and Selected features by IG.

Category	Statistics	POS	TF-IDF	Topic	N-gram	Total
Total extracted features	8	32	50	10	100	200
Selected features by IG	3	12	26	10	40	91

3.5. Classification Process.

Three classification algorithms were used in the proposed approach: Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). These algorithms were tested on all extracted features once, selected features by information gain once, and principal components generated by PCA once more. Each algorithm has its own set of hyperparameters. Under the same constraints or with the same hyperparameter values, each algorithm is applied to the three sets of features, the original features, the selected features by IG (IG Features), and the principal components generated by PCA (PCs). The hyperparameters were used to achieve the highest accuracy; this is known as fine-tuning. Table 5 shows the updated hyperparameters for all models. The results of each algorithm were tracked using the evaluation chosen metrics, which are precision, recall, accuracy, and f1-score.

Table 5: the updated hyperparameters for each model

Model	The Updated hyperparameters
SVM	kernel = ' rbf ' , max_iter= 4000 , gamma='auto'
RF	n_estimators=100 , max_depth=16
NB	var_smoothing=1e-09

4. Experimental Results and Discussion

In this section, the used evaluation metrics to evaluate our proposed methodology were presented in section 4.1, our proposed approach results were shown in section 4.2, and then the comparative analysis with the existing studies on the same dataset is discussed in 4.3.

4.1 Experimental Setup. Experiments are carried out on a laptop equipped with an Intel Core i3 CPU, 12 GB of RAM, and the 64-bit Windows 10 operating system. The NLTK and Spacy libraries were used to clean the dataset and extract our feature set. The sklearn library is used to implement the classification models (RF, SVM, and NB) and feature selection algorithms (PCA and IG) in Python. 70% of the dataset is used for training classification models, while the remaining 30% is used to evaluate the detection model's performance.

4.2 Evaluation Metrics. To evaluate our proposed methodology different evaluation metrics were used, like accuracy (Acc) Equation (4), Precision (P) Equation (6), Recall (R) Equation (5), F1-score (F1) Equation (7), and Testing time. It is based on a confusion matrix that includes information about the prediction result of each test sample. Accuracy is the correct rating rate. Precision is defined as the ratio between all cases correctly classified in the positive class versus the total number of cases classified in the positive class. In other words, the percentage of cases classified in the positive category is correct. The recall is defined as the ratio between all cases correctly classified into the positive class versus the total number of actual items of the positive class. In other words, it tells you how many positive cases were classified correctly. The closer both values are, the higher the F1 score. In the rating scales, true negative predictions (TN), true positive predictions (TP), false-negative predictions (FN), and false-positive predictions (FP). The plain rating evaluation score is the accuracy defined as follows:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Recall\ (R) = \frac{TP}{TP+FN} \tag{5}$$

$$Precision\ (P) = \frac{TP}{TP+FP} \tag{6}$$

$$F1 = 2 . \frac{Precision . Recall}{Precision + Recall} \tag{7}$$

4.3 Experimental Results. In this section, three classification methods were compared, one with original features, another with PCs, and another with IG Features. The used three classifiers SVM, RF and NB. Statistical features, TF-IDF features, n-grams features, topic features, and POS tagging features are among the five feature sets extracted. Our feature selection methods are PCA and IG. Our models were built based on a train test split methodology with a testing ratio equal to 0.3 and a training ratio equal to 0.7.

Table 6 shows the results of the proposed methodology in terms of evaluation metrics, the results of three different combinations of the features sets were compared, the first row shows the results of all feature sets that represented 200 features, the second row shows the IG features where the number of selected features eliminated to 91 features, and the third row shows the PCs where original features transformed to 67 components.

Based on the results shown in table 6, It is observed that the accuracy of all methods increases after applying feature selection methods. When applying Information Gain as a feature reduction method, the accuracy of RF and SVM algorithms increased by about 1%, and NB increased by about 8%. When applying PCA the accuracy of RF and SVM algorithms increased by about 2%, and NB increased by about 13%. As well as the testing time in all algorithms decreases significantly when applying feature selection approaches, which directly contributes to preserving people's lives before the tragedy occurs. Random Forest classifier achieves the best accuracy at 97.02% by applying PCA Feature selection in only 88.9 ms, while the accuracy was 95.75% using Random Forest but with all extracted features in 132 ms. Where the accuracy was increased by using feature selection methods. Table 6 shows the effect of applying the feature selection phase on classification accuracy.

Fig 3 shows a visualization of RF accuracy using original Features, IG Features, and PCs. Fig 4 shows a visualization of SVM accuracy using original features, IG Features, and PCs. Fig 5 shows a visualization of NB accuracy using original features, IG Features, and PCs. Where the accuracy was increased by using feature selection methods. Figs [3-5] show the effect of applying the feature selection phase on classification accuracy.

Table 6: Results of our proposed models

Methods	Features	Accuracy	F1-Score	Precision	Recall	Testing Time (ms)
RF	Original Features	95.75	95.76	95.67	95.85	132
	IG Features	96.13	96.15	96.33	95.97	95.9
	PCs	97.02	97.04	97.27	96.82	88.9
SVM	Original Features	87.44	87.52	87.77	87.27	2210
	IG Features	88.39	88.39	88.14	88.64	971
	PCs	89.42	89.59	90.68	88.52	883
NB	Original Features	76.26	77.43	81.18	74.01	31
	IG Features	83.86	84.21	85.79	82.68	20
	PCs	88.86	88.85	88.52	89.19	14

4.4 Comparative Analysis. In this section, the accuracy of our proposed approach is compared with some of the recent and existing studies in the literature. Our proposed approach has been compared with previous studies on the same dataset, there are two studies on the same dataset [40] and [41]. Table 7 summarizes the results which show that the accuracy of the proposed methodology exceeds these existing studies in the classification accuracy of suicide ideation detection. Ji et al [40] extracted five subsets of features and apply six classifiers, the best accuracy is 95.71% with the XGBoost classifier using all five feature sets. Michael et al. [41] developed a Machine learning experimental approach based on six baseline methods with combinations of three extracted feature sets and a deep learning approach. The better accuracy achieved is 93.8% by the deep learning approach with Word2vec.

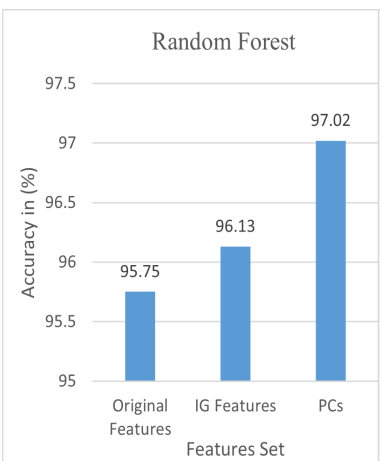


Fig 3: RF Results Analysis

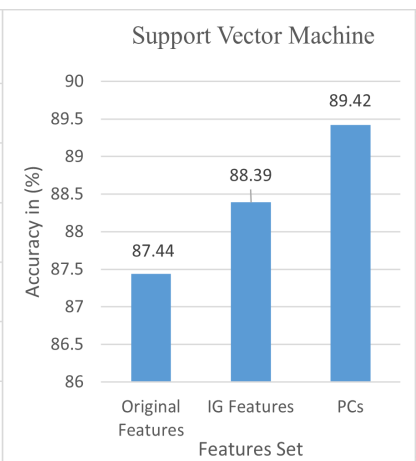


Fig 4: SVM Result Analysis

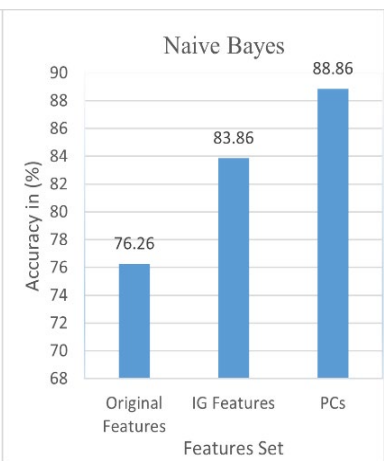


Fig 5: NB Result Analysis

Table 7: Comparison with the previous studies on the accuracy

Ref	Features	Accuracy	Model
[40]	Statistics + TF-IDF + Topic + POS + LIWC	95.71	Xgboost
[41]	Word2vec	93.8	LSTM-CNN
[41]	Statistics + TF-IDF + Bag of Words	88.3	Xgboost
Proposed	PCA	97.02	Random Forest

5. Conclusions and future work.

In modern society, early detection of suicidal thoughts remains one of the most important tasks. The huge number of negative messages, opinions and energy emitted from social media necessitates us to develop new and fast ways to detect suicidal, negative, and violent thoughts in user posts.

In our paper, a new method was designed and tested for assessing the mental health of various Reddit users. The goal of this model is reducing the rate of suicide in society by detecting suicidal intentions in individuals' social media interactions. Through NLP and machine learning techniques, fruitful analyses were presented that can improve the understanding of suicidal ideation and behaviour. This can assist to provide appropriate medical assistance and/or assistance to those in need. Our research results show that the proposed model, which is based on the feature set and importance selection, extracts more feature information and has a distinct advantage in detecting suicidal ideation. To some extent, our study fills gaps and provides new insights from previous research. This study has some advantages where context is preserved by using robust features and feature minimization, which increases model accuracy and reduces complexity.

5.1 *Theoretical Contribution*

The results of this research reveal several important theoretical contributions. First, by proposing our model based on feature combinations, this study contribute to the literature on public health and safety by enriching the idea of model construction based on full reference to previous research results. Second, through a critical review of research in the field, this study focuses on the current position and shortcomings of existing studies in the field.

5.2 Limitation and future work

This study has some limitations as well. Users' age, gender, location, or other information cannot be obtained due to Reddit's privacy settings. Another limitation is a lack of data. Data scarcity is one of the most pressing issues in current research [48], where supervised learning techniques are primarily used. They usually necessitate manual annotation. However, there is insufficient annotated data to support further research.

In future work, Other factors should be used in detecting suicide in textual data. Although informative feature sets have been extracted and a good feature selection method has been applied. However, other factors like historical data or posts of users, temporal information, and spatial information. Also applying other feature selection methods can improve the accuracy of early detection of suicidal ideation in online social media.

References

[1] A. J. Ferrari, R. E. Norman, G. Freedman, A. J. Baxter, J. E. Pirkis, M. G. Harris, A. Page, E. Carnahan, L. Degenhardt, T. Vos, and others, “The burden attributable to mental and substance use disorders as risk factors for suicide: findings from the Global Burden of Disease Study 2010,” *PloS one*, vol. 9, no. 4, p. e91936, 2014.

[2] M. Conway and D. O’Connor, “Social media, big data, and mental health: current advances and ethical implications,” *Current opinion in psychology*, vol. 9, pp. 77–82, 2016.

[3] S. Saxena, M. Funk, and D. Chisholm, “World health assembly adopts comprehensive mental health action plan 2013–2020,” *The Lancet*, vol. 381, no. 9882, pp. 1970–1971, 2013.

[4] W. H. Organization and others, “Preventing suicide: A global imperative,” 2014.

[5] M. K. Nock, G. Borges, E. J. Bromet, J. Alonso, M. Angermeyer, A. Beautrais, R. Bruffaerts, W. T. Chiu, G. De Girolamo, S. Gluzman, and others, “Cross-national prevalence and risk factors for suicidal ideation, plans and attempts,” *The British journal of psychiatry*, vol. 192, no. 2, pp. 98–105, 2008.

[6] M. A. Silver, M. Bohnert, A. T. Beck, and D. Marcus, “Relation of depression of attempted suicide and seriousness of intent,” *Archives of General Psychiatry*, vol. 25, no. 6, pp. 573–576, 1971.

[7] E. D. Klonsky and A. M. May, “Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research,” *Suicide and Life-Threatening Behavior*, vol. 44, no. 1, pp. 1–5, 2014.

[8] R. Martınez-Castaño, J. C. Pichel, and D. E. Losada, "A big data platform for real time analysis of signs of depression in social media," *International journal of environmental research and public health*, vol. 17, no. 13, p. 4752, 2020.

[9] K. Kang, C. Yoon, and E. Y. Kim, "Identifying depressive users in Twitter using multimodal analysis," in *2016 international conference on big data and smart computing (BigComp)*, 2016, pp. 231–238.

[10] S. Javadi, R. Safa, M. Azizi, and S. A. Mirroshandel, "A recommendation system for finding experts in online scientific communities," *Journal of AI and data mining*, vol. 8, no. 4, pp. 573–584, 2020.

[11] D. D. Ebert, M. Harrer, J. Apolinário-Hagen, and H. Baumeister, "Digital interventions for mental disorders: key features, efficacy, and potential for artificial intelligence applications," *Frontiers in Psychiatry: Artificial Intelligence, Precision Medicine, and Other Paradigm Shifts*, pp. 583–627, 2019.

[12] V. Venek, S. Scherer, L.-P. Morency, J. Pestian, and others, "Adolescent suicidal risk assessment in clinician-patient interaction," *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 204–215, 2017.

[13] M. Mohri and A. Rostamizadeh, "A. Talwalkar Foundations of machine learning." MIT Press, Cambridge & London, 2012.

[14] D. Cameron, G. A. Smith, R. Daniulaityte, A. P. Sheth, D. Dave, L. Chen, G. Anand, R. Carlson, K. Z. Watkins, and R. Falck, "PREDOSE: a semantic web platform for drug abuse epidemiology using social media," *Journal of biomedical informatics*, vol. 46, no. 6, pp. 985–997, 2013.

[15] P. A. Cavazos-Rehg, M. J. Krauss, S. J. Sowles, S. Connolly, C. Rosas, M. Bharadwaj, R. Grucza, and L. J. Bierut, "An analysis of depression, self-harm, and suicidal ideation content on Tumblr," *Crisis*, 2016.

[16] S. J. Sowles, M. J. Krauss, L. Gebremedhn, and P. A. Cavazos-Rehg, "I feel like I"ve hit the bottom and have no idea what to do': Supportive social networking on Reddit for individuals with a desire to quit cannabis use," *Substance abuse*, vol. 38, no. 4, pp. 477–482, 2017.

[17] S. T. Rabani, Q. R. Khan, and A. Khanday, "A Novel Approach to Predict the Level of Suicidal Ideation on Social Networks Using Machine and Ensemble Learning," *ICTACT J. SOFT Comput*, vol. 11, no. 2, pp. 2288–2293, 2021.

[18] J. Miao and L. Niu, "A survey on feature selection," *Procedia Computer Science*, vol. 91, pp. 919–926, 2016.

[19] S. Maza and M. Touahria, "Feature selection algorithms in intrusion detection system: A survey," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 10, pp. 5079–5099, 2018.

[20] K. K. Vasan and B. Surendiran, "Dimensionality reduction using principal component analysis for network intrusion detection," *Perspectives in Science*, vol. 8, pp. 510–512, 2016.

[21] M. M. Sakr, M. A. Tawfeeq, and A. B. El-Sisi, "Filter versus wrapper feature selection for network intrusion detection system," in *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2019, pp. 209–214.

[22] N. C. Jones, "Prediction and analysis of degree of suicidal ideation in online content," Massachusetts Institute of Technology, 2020.

[23] D. Sikander, M. Arvaneh, F. Amico, G. Healy, T. Ward, D. Kearney, E. Mohedano, J. Fagan, J. Yek, A. F. Smeaton, and others, "Predicting risk of suicide using resting state heart rate," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016, pp. 1–4.

[24] N. Jiang, Y. Wang, L. Sun, Y. Song, and H. Sun, "An ERP study of implicit emotion processing in depressed suicide attempters," in *2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, 2015, pp. 37–40.

[25] W.-C. Chiang, P.-H. Cheng, M.-J. Su, H.-S. Chen, S.-W. Wu, and J.-K. Lin, "Socio-health with personal mental health records: suicidal-tendency observation system on Facebook for Taiwanese adolescents and young adults," in *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, 2011, pp. 46–51.

[26] Y.-P. Huang, T. Goh, and C. L. Liew, "Hunting suicide notes in web 2.0-preliminary findings," in *Ninth IEEE international symposium on multimedia workshops (ISMW 2007)*, 2007, pp. 517–521.

[27] K. D. Varathan and N. Talib, "Suicide detection system based on Twitter," in *2014 Science and information conference*, 2014, pp. 785–788.

[28] J. Jashinsky, S. H. Burton, C. L. Hanson, J. West, C. Giraud-Carrier, M. D. Barnes, and T. Argyle, "Tracking suicide risk factors through Twitter in the US," *Crisis*, 2014.

[29] W. Wang, L. Chen, M. Tan, S. Wang, and A. P. Sheth, "Discovering fine-grained sentiment in suicide notes," *Biomedical informatics insights*, vol. 5, p. BII–S8963, 2012.

[30] S. J. Cash, M. Thelwall, S. N. Peck, J. Z. Ferrell, and J. A. Bridge, "Adolescent suicide statements on MySpace," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 3, pp. 166–174, 2013.

[31] A. Shepherd, C. Sanders, M. Doyle, and J. Shaw, "Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation," *BMC psychiatry*, vol. 15, pp. 1–9, 2015.

[32] H.-Y. Huang and M. Bashir, "Online community and suicide prevention: investigating the linguistic cues and reply bias," in *Proceedings of the conference on human factors in computing systems*, 2016.

[33] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "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*, 2016, pp. 2098–2110.

[34] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on twitter," *Computer communications*, vol. 73, pp. 291–300, 2016.

[35] M. Kumar, M. Dredze, G. Coppersmith, and M. De Choudhury, "Detecting changes in suicide content manifested in social media following celebrity suicides," in *Proceedings of the 26th ACM conference on Hypertext & Social Media*, 2015, pp. 85–94.

[36] 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 mental health*, vol. 3, no. 2, p. e4822, 2016.

[37] H. Sueki, "The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan," *Journal of affective disorders*, vol. 170, pp. 155–160, 2015.

[38] B. O'dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Interventions*, vol. 2, no. 2, pp. 183–188, 2015.

[39] E. Okhapkina, V. Okhapkin, and O. Kazarin, "Adaptation of information retrieval methods for identifying of destructive informational influence in social networks," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 2017, pp. 87–92.

[40] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," *Complexity*, vol. 2018, 2018.

[41] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," *Algorithms*, vol. 13, no. 1, p. 7, 2019.

[42] S. Renjith, A. Abraham, S. B. Jyothi, L. Chandran, and J. Thomson, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 9564–9575, 2022.

[43] J. Liu, M. Shi, and H. Jiang, "Detecting suicidal ideation in social media: An ensemble method based on feature fusion," *International journal of environmental research and public health*, vol. 19, no. 13, p. 8197, 2022.

[44] S. Maruf, K. Javed, and H. A. Babri, "Improving text classification performance with random forests-based feature selection," *Arabian Journal for Science and Engineering*, vol. 41, pp. 951–964, 2016.

[45] D. Kumar and others, "Feature extraction and selection of kidney ultrasound images using GLCM and PCA," *Procedia Computer Science*, vol. 167, pp. 1722–1731, 2020.

[46] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Technical Review*, vol. 38, no. 4, pp. 377–396, 2021.

[47] E. O. Omuya, G. O. Okeyo, and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," *Expert Systems with Applications*, vol. 174, p. 114765, 2021.

[48] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal ideation detection: A review of machine learning methods and applications," IEEE Transactions on Computational Social Systems, vol. 8, no. 1, pp. 214–226, 2020.