A Novel Approach for Communication-related to suicidal detection on Twitter using multi-class data

Rajesh Kumar[1,*], N. Venkatram[1]

[1]Department of CSE, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, AP, India
Emails: rajthalopo@gmail.com; venkatram@kluniversity.in

Abstract

Suicide is a significant issue for public health worldwide since suicide is not something that happens randomly but is influenced by social and environmental variables as well. At the same time, effective early diagnosis and treatment may lead to several positive health and behavioural results. Suicide persists undiagnosed and untreated for many reasons, including denial of sickness and cultural and social disgrace. Through the ubiquity of social media, by expressing opinions, thoughts and everyday struggles with mental health on social media, millions of people are sharing their online identity. As opposed to typical retrospective research that uses self-reported surveys and questionnaires, this study assesses the validity of identifying suicidal symptoms using Twitter tweets that were gathered over more than a year, using a variety of online web-blogging sites as points of reference. For recognizing tweets expressing suicidal thoughts, three sets of characteristics are employed for training the dataset employing base and ensemble classifiers. The Rotation Forest (RF) approach is the preferred baseline, and the Maximum Probability Voting Decision approach is used in seven different labelled classes relating to suicide communication and class demonstrating suicidal thoughts. With the suicidal ideation class scoring 0.76 and the suicidal contents for all seven classes scoring 0.82, this revised model was able to attain an F-measure. To increase awareness of the vocabulary made use of on Twitter to express suicidal thoughts, the findings are summarized by highlighting the predictive principal component of suicide communication in classrooms.

Keywords: Suicide; Public health; Social disgrace; Twitter tweets; Maximum probability; principle component, Rotation Forest; suicidal communication; decision method.

1. Introduction

According to [1], there is data associating social reporting on suicide instances with suicidal behaviour, and many worries concerning how social media affects suicidal thoughts and spreads among vulnerable subjects have been brought up [2]. The emergence of popular microblogging and social networking sites, like Twitter, Tumblr, and Facebook (together called "Social Media"), has been discussed as a potential new form of highly dynamic, the most advanced platform for group communication. It could result in the illness of suicidal ideation, similar to how media coverage is broadcast.

Social science and medical studies, particularly with younger generations, have examined the influence that suicide via the World Wide Web (WWW) has on this sensitive topic [3]. [4] The research performed quantitative research by questioning adults involved in suicidal activities and websites depicting the theme of suicide. [5] It states the conducted online research on different websites for words exhibiting suicidal terms and related methods of suicide. They provided a qualitative study of the tool they found. They decided that, even though anti-suicide and neutral websites were most commonly supportive, websites and pro-suicide forums promoting suicidal

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

behaviour were also available, indicating that more internet-based preventive strategies are required. Working on this, [6] the reviewed online literature on suicide prevention and intervention, more attention needs to be paid to creating and assessing online preventive methods because there is a shortage of empirical research on the subject. [7] investigated the impact of suicide content on social media, specifically Facebook, on suicidal behaviour. There arises a question to what level suicide messages on social media sites can directly cause suicide to copycat. They communicate that suicide and the consequences of social media require further study and evaluation.

In [8] focused on web-based reporting of suicide through display boards and newsletters, chatrooms [9], newsgroups [10] and web forums [11]. All these are primarily qualitative analyses where online web-related suicide studies use quantitative data; these still only use human interpretation, which is challenging to do on a large scale. A limited subset of suicide communication studies employ statistical methodologies. The frequency of tweets that can suggest suicidal ideation and intent is positively correlated with the suicide rate, according to research [12]. The identification of suicide language on Twitter is the subject of an emerging corpus of scholarship, according to J. F. Gunn and [13]. However, very few academics employ categorization to identify language that invariably reflects suicidal ideas and to differentiate suicidal language from other suicide-related communication forms, like reporting suicide and creating awareness. The variation is a prerequisite for finding the symptoms and characteristics of suicidal intent on Twitter. In [14] and [32], the proposed research on depression and other emotive states is shared through social media (Twitter). Suicidal thoughts comprise sensitive contents and probable signs of depression, but in this paper, suicidal ideation and depression are synonyms. Helping suicidal media users begins with an understanding of how suicidal thought is expressed [33-36]. According to a recent study, people are more likely to turn to unprofessional resources—like social media (such as Twitter)—for assistance rather than taking a chance and seeking out formal medical care.

This work aims to contribute to the topic of Twitter suicide by (i) improving the definition of suicidal traits by manually generating a dataset. (ii) establishing a benchmark and experimenting with a set of outcomes for suicidal ideation using classification techniques, and (iii) building a machine classifier that is fit for recognizing depressive language such as Flippant references to suicide, suicidal ideation and perception and information about suicide. As it emphasizes primarily the risk to human security and safety than it does on suicide references, the last addition is particularly pertinent for estimating the proper amount of gloomy language on social networking platforms. This paper focuses on examining the data or tweets collected from Twitter; the data are classified into seven different categories and applying Machine learning (ML) classification models to distinguish suicidal tweets to predict and analyze suicidal ideation to help clarify the language used by Twitter users.

2. Related works

In recent studies, [15] analyzed clinical conversation and suicide contents to generate classification and to recognize emotions and topics stated by individuals who have attempted suicide [16], [17], [18], [19], and [20]. These papers aimed to categorize text in terms of paragraphs or sentences, which implies short sentences were also included and posted on Twitter. However, individuals who had accepted suicide and harmed themselves have written suicidal notes. At the same time, existing research focuses on finding suicidal ideation before suicide, which may vary from the suicidal language used in the notes of suicide. In addition, manually written notes are not inhibited by string size at the sentence level. On Twitter, the user is forced to use only informal and short language distinct from the usual expressive way of feeling on the wrapper. Eventually, Twitter data content is noisy and encompasses a wide variety of topics varying over time. These characteristics make it more complex to classify the suicidal idea than to record feelings and pre-suicide thoughts on suicidal notes.

Few researchers have experimented with how suicidal ideation is communicated on Twitter by comparing it with national mortality rates. For instance, studies done in the US [21] and [22] have searched for a significant relationship between the number of suicide posts on social media and the number of reported suicide cases. Suicide contents that are related to ideas like depression and suicide [23] or specifically to particular risk concerns [24] were identified using a set of keywords. To find a behavioural development between suicidal ideation and psychological difficulties, [25] took into consideration websites like Suicidewatch /r on Reddit and presented subject analysis and language features. As a result, the threat of encouraging messages in online support is underlined. [26] examined the tweets which contain suicidal content or thoughts by analyzing their interactions and relationships in social networking rather than the tweets posted. These adjustments or changes may cause effects

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

due to different factors leading to suicide among celebrities [27]. In [28], a statistical relationship has been identified between suicide rates and more posts associated with suicidal celebrities in Japan. A recent suicide victim was studied in [29], who saw sudden shifts in the subject's perspective on dying and an increase in positive sentiments (albeit not statistically significant) in the twenty-four hours before the person's death on Twitter. Given that this was a study of a person who had tried to make rhyme posts (possibly in different suicide languages to attain this), the author recommended extensive studies on Twitter messages. Parts of speech (POS), expressive phrases, and cognitive processes are among the ideas that can be categorized using the Linguistic Inquiry and Word Count (LIWC) method, according to the research carried out [30]. The LIWC sample technique was also utilized in [31] to identify "sad" tweets or messages classified to the usual distress scale with a best-case accuracy of roughly 65%.

2. Data Collection

Terms of the lexicon are not created by compiling a list of words with suicidal content but rather by gathering information from web forums, microblogs, and known suicide blogs. To debate suicidal ideas for prevention and support, the user-generated tweets are first collected from four websites recommended by experts in this field. These websites might have required sections (like "http://www.enotalone.com and http://www.experienceproject.com"), or they might include content (like "http://www.takethislife.com and http://www.recoveryourlife.com") that encourages suicide. Information is also obtained from a microblogging platform (e.g., "https://www.tumblr.com") where posts exhibiting suicidal conduct are archived, including text including the term "suicide."

Table 1: First 20 Tri-grams and five-grams

TF-IDF	Tri-Grams	TF-IDF	Five-Grams
269.93	"To kill me."	33.89278	"Want to die right now?"
226.35	"Want to die?"	29.63562	"To take my own life."
91.74	"Want to kill?"	28.50259	"It is not worth it anymore."
78.17	"To commit suicide."	21.61567	"Have nothing to live for."
65.63	"To end it."	21.61567	"I want to kill myself."
60.19	"Cannot live."	21.61567	"I no longer want to live."
59.13	"I hate myself."	21.61567	"Want to stay here no longer?"
55.45	"I am tired of."	21.4643	“Myself hate my life hate”
54.80	"End my life."	20.45271	“Want it all to end”
48.46	"End it all."	20.45271	"Want it to be over?"
38.96	"Kill me and"	20.45271	"Fall asleep and never wake."
35.88	“Take my own”	18.62250	"Wish I could just fall asleep."

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

33.81	"To live anymore."	17.93278	"I just want to die so bad." "Want everything to end?"
32.77	“My death would”	14.27610	"Want to end it all?"
32.60	"Kill me, I"	14.27610	"I am sorry that I am leaving."
30.71	"About killing myself."	14.27610	"Rather die, it is not worth it."
29.72	“Killing myself I”	14.27610	"So why should we continue living?"
29.14	"Never wake up."	14.27610	"Fuck trying to live normally."
27.26	“Stop the pain”	14.27610	“To commit suicide within few."
27.24	"Kill me, right?"	14.27610	"Want to avoid living a defeated life?"
26.88	“Point in living”	14.27610	"Put an end to this."
26.85	“Thoughts of suicide”	14.27610	"And pain anymore just can."
25.62	“Have nothing to”	14.27610	"Bad am worthless what."
25.30	"Worth it anymore."	14.27610	"Been self-harming for years."
23.85	"Wanted to die."	14.17610	"Life is this miserable just."

From the listed websites, more than 1000 posts are selected randomly from Tumblr; the post with more extended text size is avoided by text greater than 5% of the average text length. Find terms with a rise in suicidal ideation and a decrease in term frequency with postings that are not suicidal using term frequency/inverse document frequency (TF-IDF). This approach examines n-grams ranging from 1 to 5 to remove duplicate phrases and non-suicidal messages. Examples of relevant n-grams generated by TF-IDF are shown in Table 1.

Table 2: Keyword and Search Terms

"Die for nothing."	“Suicidal Thoughts”	"Cease to exist."	“Breathe one’s last”
"Take from me."	"Still, I live."	“All over”	"Why should I continue living?"
“Hang all over."	“Suicide myself”	“End up”	“Killing me”

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

"Be no more."	“Hate myself”	"Asleep and never wake."	"Have to die."
"Die me"	"Take my sole."	"End with this."	"To hang out."
“To demise”	"I do not want to survive."	“My death would”	"Feeling tired of being by myself alone."
"Life is too hard."	"To hurt me."	"Want to be alive anymore?"	"I would rather not wake up."
“Die in my sleep”	"Want to be around anymore?"	"End it all."	"Want it to be over?"
“Thoughts of suicide”	"Life is so meaningless."	"Not want to be alive."	“Want to be gone”
"Just want to sleep forever."	“Nothing to live for”	"Want to be here anymore?"	"What is wrong with me?"
"Could just fall asleep."	"To end this nightmare."	“Hate my life”	"Put an end to this."
"Cannot do this anymore."	"My life is pointless."	“Ready to die”	"My life is not worth"
"Take my own life."	"Do not want to go on"	"Wanted to die."	"I am leaving now."
"Want to die?"	"End my life."	"I am worthless Suicidal"	“Wanting to kill yourself’
"Is not worth living Suicide."	"My life consists of nothing."	“Just want to give up."	"Take it anymore."
"Kill me."	"To live anymore."	“Killing myself’	"Ending it all."
“Tired of being alone”	"Do not want to try anymore."	"End this pain."	“Point in living”
"Do not want to be here."	"Want to disappear?"	“Stop the pain”	“Want to end it?"
"Life is worthless."	"Want to be dead?"	"I am drowning."	"Do not want to exist."

Since Twitter contains many unrelated terms, the keywords generated would only be helpful to search in Twitter post collections sometimes. The TF-IDF terms were further examined to generate additional keywords to identify suicidal communication on Twitter, as shown in Table 2. These Keywords are

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

used as search terms in Twitter content, and data have been collected from Twitter API for more than a year starting from 1 February 2018, resulting in more than 2 million posts of the dataset. To characterize how people communicate about suicidal issues, seven different labelled classes are considered, as shown in Table 3.

Table 3: Suicide Communication Types

Class	Description
Cl	Evidence of suicide attempts
C2	Support or information
C3	Suicide flippant reference
C4	Campaign or fight (petitions etc.)
C5	Condolence or memorial
C6	Suicide reporting
C7	None of these

4. Feature selection

To train and test different classifiers, tweets are used to determine originality and to distinguish between suicide and another form of suicidal communication, even a careless reference to suicide. From tweets, the following three sets of features are extracted:

- ✓ Phrases and words used more commonly are examples of structural language elements and additional sentence lexical aspects like Parts of Speech (POS). These are standard features utilized in activities related to text mining. POS also captures references to others and self—these terms have been linked to suicidal reporting in past studies.
- ✓ Terms with characteristics that convey mood, affective, and emotional levels are employed in the text.
- ✓ Regardless of the unique emotional complexity nature, these were integrated. Emotions like anger, fear and general violence in suicidal communication are especially prominent.

In social media posts, features describing characteristic language within limited characters are stated in short and informal text. All of them were taken from Tumblr annotated posts to try and incorporate terminology from Twitter media that might not be picked up by standard text mining algorithms [37-38].

4.1. Feature selection-1

The requirement for the first feature set and half of the second collection were derived as applied in (I. Spasic et al., 2012), published related to the depressive analysis of suicidal notes. It is considered as Set1 features based on the following:

- POS: The log-linear POS tagger (https://nlp.stanford.edu/software/tagger.shtml) adds a POS mark to any word on Twitter. Contains nouns (categorized as "singular," "plural," and "proper nouns"), verbs (including tenses like "present," "past," and "present participle"), references to first- and third-person objects, adjectives, and adverbs ("comparative" and "superlative"), pronouns ("personal" and "possessive"), and additional labels for "determiners," "conjunctions," "symbols," "cardinal number," and "interjections." One feature that is taken into consideration for POS tagging is the frequency of each tweet.

- Structural properties: First, the inclusion of contradictory sentences is considered (total count in number), the use of pronouns in connection with the first person ("singular" or "plural")

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

and the inclusion of URL as external features in a text or by any symbols mentioned (representing reply or retweets) are examined.

- Lexical features: such characteristics illustrate classes like religion, home, sociology, psychology, etc. which are collected from Wordnet ("http://sentiwordnet.isti.cnr.it").

Emotional lexical features: associated with domains describing "affective" ideas. It includes terms that reflect emotional circumstances, moods, or expressive responses like anger, joy, sadness, grief, enthusiasm, love, hate, happiness and surprise; more prevalent subcategories include aggression, friendliness, nervousness, and irritability, and negative fear converses like worry to positive-negative, negative anxiety, positive fear, self-confidence, self-consciousness, and self-indulgence. All these are suitable for investigation for this research related to specific learning.

- Sentimental score: SentiWordNet ("http://sentiwordnet.isti.cnr.it") is used for each word in a sentence with positive and negative scores ranging between zero and one. By summation, all words for a sentence are considered as features.

- Language Words: The training model contains only frequently used phrases and n-grams (the top 100 terms), such as "unigrams," "bigrams," and "trigrams."

Keyword and phrases: Extracting 76 keywords from various site types is used to pre-filter any search phrases (e.g., "kill me," "want to end it," "want to die," "want to disappear," "do not want to try anymore," "asleep and never wake," "end it all," "my life is pointless," "to live anymore," etc.). to represent any suicidal thoughts in the tweet, a binary feature is created by merging all of the individual search term features.

4.2. Feature Selection 2

By considering the emotional and psychological expression of suicidal intent, a second selection of features is explored by applying the text evaluation tool, Linguistic Inquiry and Word Count (LIWC) to derive more precise labels describing feelings and affective emotions within the Twitter text. It provides a broader overview of terms that may be more applicable to the specific language of depressive emotions to be expected and present in suicidal ideation. Examples of such terms include senses (hearing, feeling, and seeing), money, health, religion, achievement, and three other classes of terms related to "cognitive mechanism," "social words," and "affect." The terms associated with certainty, insight, inhibition, casual, inclusive, and exclusive have been divided into sections that reflect friends, family, and people: anxiety, sadness, rage, and positive and negative emotions. Features subset related to these topics was used. However, a broad level of features has been introduced to enable the research to differentiate between depression and other types of suicidal communication (for example, support, grief, and reporting).

4.3. Feature selection 3

Social media content, such as informal spelling, grammar, and short data types, may be noisy. A set of pattern-matching and regular expression (RegEx) rules developed from the compilation of suicidal posts gathered from the Tumblr site is referred to as a feature set. Initiated language from brief, casual writing alluded to six categories associated with suicide; this is annotated as part of the human comments process that was previously carried out. To support this classifier, the example expressions for each class are numbered from 1-6 as follows, including:

Terms like "call/offer for/of help" and specific terms like "shut" (example: shutting down internet) and "stop" (example: intimidating) are referred to as ". +(\offer|\call|\ask|\need). +\help. +".

**(\cutting|\sui|\depress) |\sad|\bad\|these). + (\feel|\thoughts). +" to indicate words such as "suicidal / cutting / bad / these …. thoughts / feelings"; ". +\w\wan+ +d[ie]. +" for the emotions such as "want/wanted/wanting to die'; ". +\ends. + (\life|\it|\all). +" to denote words with "end/ending it all" and "end my life"; similarly, ".+(take+\don.+|\can).+(|\alive|\cop|\anymo|\live|\go).+" encompasses many expressions like "cannot take anymore," "cannot/do not want to live/cope anymore," "do not want to be alive," "cannot take it anymore," and "cannot go on." Additionally, number of n-grams and single words like "trigger alert,", "we", "eating `death', disorder", "anxiety", "self-harm" and "pain" are included.

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

About expressions like "talk/speak to someone/someone" and terms like "blog", "web", "advice", and "health. " +(\speak|\talk). +\to. +(\any|\some|\one).

".+(\throw\kill\hat)" refers to phrases like "kill/killing /hate me," ".+(\f***.+" refers to profanity like "f**k/ f**king," ".+ (\girl\boy).+(\friends)" refers to languages with "boyfriend," "girlfriend," and ".+(\like)\.+(\just).+" refers to phrases like "just"... "like" "sexual reference". Further, some terms related to topics like "job", "school", and "college" were also included, as they reflect situations that are more conducive to flippant language apart from serious expressions of depression and suicidal ideation.

". +(\take|\took). +\his\her\him\own. +\life. +" phrases include "took/taken his/her own life" and terms meaning "overdose," "hanged," and "hanging," which convey suicidal ideas.

Phrases like "miss/missing you/her/him" and ". +(\die|\kill|\comm). + (DD|MM |YY). +" are used to indicate precise time intervals.

The regular RegEx used in the fourth class denoting flippancy was found (and vice versa) among those related to the second class. However, it is confirmed to associate regular expression to any of these single classes depending on the annotated Twitter tweets. For example, keywords like "kill me" or "hate me" were often related to flippant Twitter posts, where words such as "want to end it" and "wanted to die" were also included in text exhibiting evidence of suicidal thoughts.

4.4. Data-driven process

The three features are combined in a combination like feature1 and feature2 as feature4, feature1 and feature as feature5, feature2 and feature3 as feature6 and the seventh set of features is referred to as a collective set by performing a union of all three features as discussed in section 4. To reduce the size of a dataset, Principal Component Analysis (PCA) is utilized to transform all features into a set of principal components, possibly non-correlated characteristics. It is done because each Twitter tweet has a large set of features linked, and there is a chance that these features will collide. Tweet text is integrated as a function set for all experiments.

In this process, n-grams of size 1-gram to 5-grams are used to transform the tweets into a word vector; the words in the order increments like "100, 300, 500,1000, 1500 and 2000" are retained. From these words, the optimal performance of 1-gram to 3-gram words is retained for this research.

4.4.1. Disadvantages of Existing System

The existing work was examined with a collection of 6 months of data by considering all tweet posts without suicide-related communication. So, to perform a practical suicidal analysis, data with suicidal thoughts for an extended period is required.

5. Methodology

Methodologies concerning baseline and ensemble classifications are performed and explained in the below sections

5.1. Baseline Classification Analysis

5.1.1 Rule based classifier:

The rule set is predicated on the record classification in a rule-based approach. Equation 1 depicts the normal disjunctive form of a set of rules.

$$\beta = (\alpha 1 V \alpha 2 V .....V \alpha_k) \quad (1)$$

Where β- denotes the set rule

αi- denotes rules for classification

the αi classification rule is

$$a_i : (Cond) \rightarrow x_i \quad (2)$$

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

where Cond- denotes antecedent, x_i– denotes consequent

The first step in creating a rule-based classifier is to extract rule set β using equation 1. This article uses a rule-based approach called Decision Tree Majority (DTM). It has two components: body and schema. The body is a named instance set from the space identified by the characteristics in the schema; schema is a collection of feature sets added to the table. The feature-selection model selects the functions or features used in the schema. DTM model seeks an accurate match for unlabeled classes by applying different features in the schema. If the unlabeled classes match the content, then it returns the class label of the majority class. Otherwise, it returns the DTM majority class. While returning the DTM majority class label, the information gain can be biased, with the root node containing the more considerable value.

5.1.2. Naive Bayes classifier

It is used as a probabilistic approach. To perform prediction, the following steps are followed: first, all the data are converted into a frequency table; the second step is creating the probability table; the last step is calculating the posterior probability for every class for prediction, and the class with maximum probability is the predicted output.

5.1.3. Support Vector Machine

SVM aims to split the dataset into different classes to determine the Maximum marginal Hyper-plane (MMH) for prediction.

5.2. Ensemble Classification analysis

A group of classifiers arranged in a tree form comprise the proposed rotation forest (RF) classification model. Let $y = \{y_1, y_2, \ldots\ldots\ldots y_n\}$ be the data point that feature n describes, and let Z be a $N \times n$ matrix-shaped dataset. X is a vector representing class labels, $X = \{x_1, x_2, \ldots\ldots\ldots x_n\}^T$ where x_j the ensemble classifier denotes the value from the class label set $\{\theta_1, \theta_2 \ldots\ldots\ldots, \theta_c\}$ $\mu_1, \ldots\ldots\ldots, \mu_L$ and F_s the feature set. With the ensemble model, the L value should be considered in advance. Hence, all the classifier is parallelly trained with RF and bagging approach. To construct μ_i a classifier, the following steps are followed.

W subsets are randomly selected from the feature set F_s, where W stands for a parameter. In this case, the subsets may intersect or not. If W is a factor of n, then each subset of features contains $O = n / W$ a feature.

$F_{s_{ij}}$ denote the jth subset feature of the classifier μ_i, the non-empty subset is selected randomly for every subset of a class and a bootstrap sample is drawn for every object of data count with the size of 70 percent. PCA is executed by using only O features in $F_{s_{ij}}$ the selected Z subset. The coefficient of PC $l^{(1)}{}_{i,1}, l^{(2)}{}_{i,1}, \ldots\ldots, l^{(z1)}{}_{i,1}$ is stored for each size $O \times 1$. Some possible eigenvalues may exist as 0, leading to fewer O vectors $O_j \leq O$. In an attempt to avoid similar coefficients when the same subset of features is chosen for several classifiers, PCA is performed on the class subset instead of the complete set.

The obtained vector can be organized in a sparse "Rotation "matrix T_{m_i} as shown in Eq. (3). The rotational sparse matrix will have dimensions $n \times \sum_j O_j$ μ_i. First, the column T_{m_i} feature is rearranged to evaluate the classifier training set corresponding to the original feature. The rotated matrix or rearranged matrix can be denoted by $T_{m_i}{}^l$ (size $P \times n$). Then, the classifier training set μ_i $ZT_{m_i}{}^a$ is.

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

$$Tm_i = \begin{bmatrix} l^{(1)}{}_{i,1}, l^{(2)}{}_{i,1}, ..., l^{(z1)}{}_{i,1} & [0] & [0] & [0] \\ [0] & l^{(1)}{}_{i,1}, l^{(2)}{}_{i,1}, ..., l^{(z1)}{}_{i,1} & [0] & [0] \\ [0] & [0] & l^{(1)}{}_{i,1}, l^{(2)}{}_{i,1}, ..., l^{(z1)}{}_{i,1} & [0] \\ . & . & . & . \\ [0] & [0] & [0] & l^{(1)}{}_{i,1}, l^{(2)}{}_{i,1}, ..., l^{(z1)}{}_{i,1} \end{bmatrix} \quad (3)$$

Because decision trees are sensitive to feature axes' rotation and can still be extremely precise, they are regarded as the base classifier in the suggested model. By using PCA, it is observed that only a small number of components can be retained; therefore, modifications to some discriminatory components may result in a small variance and be discarded; in this case, however, all components are retained, and this does not imply that feature extraction can be classified as an easy task. However, by performing rotations, if the performance outcome for discriminatory directions is less, it is diversifying heuristic. The proposed diversity in this approach is obtained from the feature subset difference. By rotational heuristic, it can be in a total $\phi = n! / W!(O!)^w$, w subset with a size O partition feature, with each subset raised to a classifier (n=WO). If there are L classifiers in an ensemble, the probability of different classifier $P = \phi / (\phi - L)!\phi^L$ s is. Extra randomization is required for the ensemble to have a different classifier. Applying PCA to the sample boost trap from X will accomplish this by randomly selecting Z or by a random subset of Z. In this proposed model, both heuristics are used. Hence, PCA is applied on all new datasets by using subset features to train the model instead of deriving PC in a single pass from all features, which would use a more extensive range of PCs to reduce the false-negative count results. So, this process is repeated with a proposed RF ensemble model.

$$\hat{L} = \arg\max(class_1, class_2, ..., class_n) \quad (4)$$

$$\theta_j(y) = \frac{1}{L}\sum_{i=1}^{L} f_{i,j}(yT_{m_i}), \quad j = 1, ..., c \quad (5)$$

The algorithm of the proposed model is as follows:

Algorithm 1:

Phase: Training

Given

- Z: training set object (N * n Matrix)
- X: training set label (N * 1 Matrix)
- L: Ensemble classifier
- W: Subset Numbers
- $\{\theta_1, \theta_2, ..., \theta_n\}$ class label set

For j=1……L

- Initialize the rotation matrix $T_{m_i}{}^{l}$

  -to split feature set F_s into $F_{i,j}$ subsets
  For k=1…...W
  - $Z_{i,j}$ denote dataset *Z* for $F_{i,j}$ the feature.
  - from random class, eliminate $Z_{i,j}$

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

- select sample(bootstrap) of size 70% from $Z_{i,j}$ the concerning $Z_{i,j}$ object, $Z^{1}_{i,j}$ denote new set.
- Run PCA on the dataset $Z^{1}_{i,j}$ to attain matrix coefficient $M_{i,j}$

- organize $M_{i,j}$ for k=1……W in the matrix after rotation T_{m_i} in Eq. (1)

- construct the rotation matrix $T_{m_i}{}^{l}$ by reordering the matrix column T_{m_i} to match $F_{i,j}$ feature.

- build μ_i a classifier by using $(Z T_{i,j}{}^{l}, X)$.

Phase: classification

For given y, $f_{i,j}(yT_{m_i}{}^{l})$ represents the probability given by μ_i classifier from θ_j class, each class's confidence θ_j is calculated using Eq. (5).

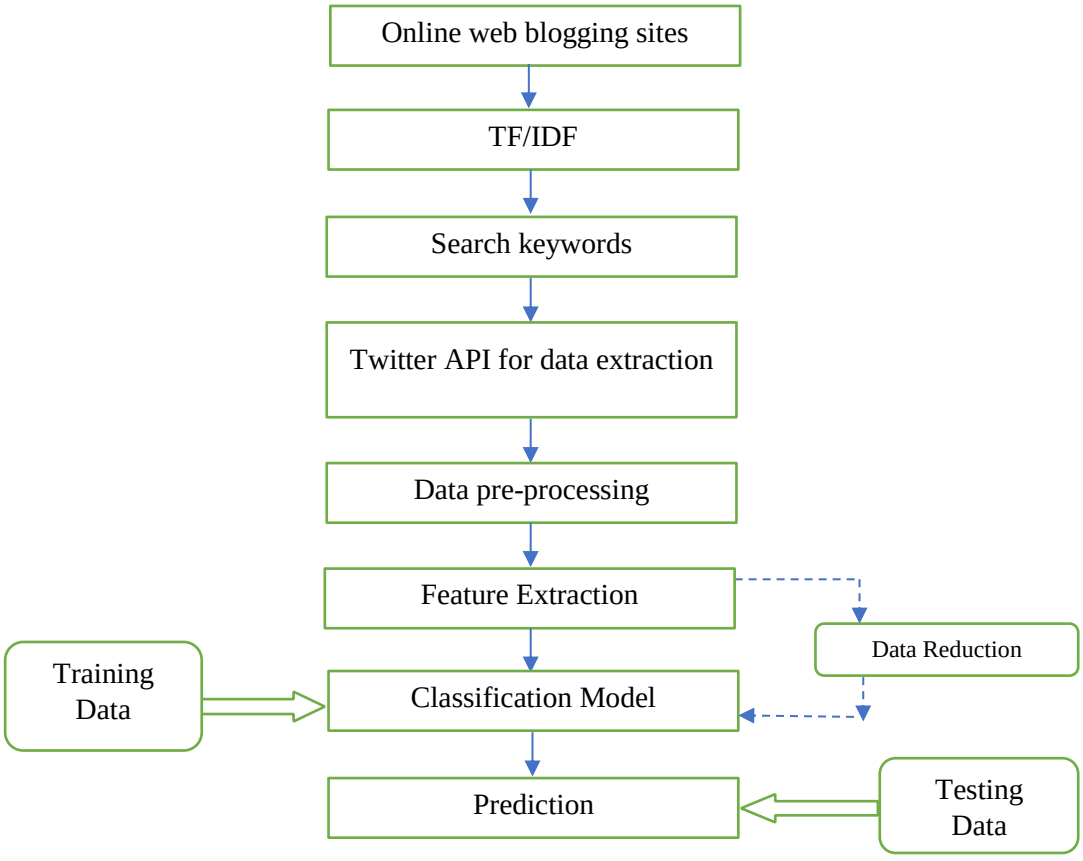

Figure 1: Proposed Model

An ensemble classifier can integrate several collective baseline models like a meta classifier. All the baseline models are incorporated to experiment with how the RF principle can enhance this classifier. Since the initial outcome exhibited differing performance among classifiers- for example, NB produced the lowest false-positive (FP) numbers with feature1 and feature3, using both features, SVM produced the lowest false-negative (FN)- a second meta classifier is integrated into the RF, using the technique that is comparable to the voting principle, it applied Eq (4) to assign and distribute the label

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

with the most significant probability across all baseline classifiers for new cases. The proposed model is shown in Figure 1. DT, SVM and NB classifiers were incorporated as an ensemble into the RF model, and the classifier with maximum probability as output for new instances was selected for decision-making. Two RF-method experiments- one is SVM and NB classifier, and another is with all three base classifiers. In Table 5, there is a notable performance difference with suicidal ideation using the DT classifier. Thus, it was removed, and the ensemble model performed better and reported for SVM and NB results combinations.

6. Results analysis and Comparisons

For evaluation, a 10-fold cross-validation method is used; this method trains 70% of training data and tests for the remaining 30%. The findings are acquired by computing the mean for each model once the iterations are finished. In this part, the results are presented on two levels. The results for every base classifier- SVM, DT and NB are presented in Table 4 for "Weighted average results for all classes" and Table 5 for "suicidal ideation results", with different feature set results in each row. The last column presents the proposed RF classifier result. The evaluation trailed standard criteria for precision (P) determining FP, Recall (R) as FN and harmonized mean for F-measure (F). Tables 4 and 5 represent the best scores in Bold and recall and precision as italic for every feature.

Table 4: Classification outcomes for every class

Classifier \ Features		**NB**	**SVM**	**DT**	**Proposed model**
Feature1	**P**	*0.794*	0.792	0.735	0.677
	R	0.781	*0.789*	0.741	0.669
	F	0.787	*0.790*	0.738	0.672
Feature2	**P**	0.783	*0.798*	0.720	0.808
	R	0.767	*0.794*	0.722	0.807
	F	0.775	*0.799*	0.720	0.807
Feature3	**P**	*0.794*	0.790	0.738	0.763
	R	0.779	0.786	0.742	0.701
	F	*0.789*	0.787	0.740	0.731
Feature1 and two combined as Feature4	**P**	0.788	0.786	0.720	0.721
	R	0.774	*0.790*	0.731	0.773
	F	0.781	0.788	0.725	0.746
Feature1 and three combined as Feature5	**P**	0.796	0.790	0.724	0.780
	R	*0.787*	0.780	0.735	0.772
	F	*0.791*	0.785	0.729	0.776
Feature 2 and 3 combined as Feature6	**P**	0.788	0.782	0.725	0.801
	R	0.773	0.788	0.730	0.799
	F	0.780	0.785	0.728	0.780

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

Feature 1,2,3 as Combined	**P**	0.774	0.795	0.722	**0.813**
	R	0.759	0.789	0.717	**0.830**
	F	0.766	0.792	0.719	**0.821**

For feature set- features 1 to 6, three base classifiers performed equally across all levels, with SVM performing better in most cases than NB and DT performing less, as shown in Table 4. NB produced the best precision score in two cases out of 3, and SVM produced the best for all three 3 cases, helping to examine with ensemble classifiers. Implementing the Dimensionality reduction method for all features resulted in decreased performance, as shown in Tables 4 and 5 in combined features. However, a performance increase was observed using the proposed RF with Maximum Voting meta classifier from P = 0.796, R = 0.794, F = 0.799 to P = 0.813, R = 0.830, and F = 0.821 among all classes. It resulted from splitting the training model into smaller samples and extracting the PC for each sample, which broadened the component range while preserving complexity. Table 5 illustrates that baseline classifiers perform worse across the board the deeper into the eagerness class relative to suicidal classes. Table 6's confusion matrix for the exceptional performance shows a significant amount of ambiguity between the C1 (class1)-suicidal ideation and the C2 (class2)-flippant relation to suicide. Selecting this will always be difficult because of the subjective nature of the undertaking and the difficulties experienced by human annotators. Irony and sarcasm are significant issues in text classification that need to be addressed. It is mainly because both severe and absurd cases employ the same word. The SVM Base classifier performed significantly better than the suggested RF model, with a precision score 0.725. However, SVM had less score when compared to other classification methods and RF for its performance. It is in context with proposed work with better performance compared to the existing field of research. However, applying PCA for a subset of trained data, the RF model produced better results when combined with all sets of features related to the suicidal class.

Table 5: Classification results for all suicidal classes

Classifier \ Features		**NB**	**SVM**	**DT**	**Proposed model**
Feature1	**P**	*0.515*	*0.725*	*0.464*	*0.589*
	R	*0.732*	*0.564*	*0.410*	*0.475*
	F	*0.605*	*0.633*	*0.435*	*0.526*
Feature2	**P**	*0.492*	*0.652*	*0.397*	*0.584*
	R	*0.706*	*0.577*	*0.372*	*0.421*
	F	*0.580*	*0.612*	*0.384*	*0.489*
Feature3	**P**	*0.506*	*0.647*	*0.530*	*0.618*
	R	*0.604*	*0.564*	*0.449*	*0.448*
	F	*0.551*	*0.603*	*0.486*	*0.519*
Feature1 and two combined as Feature4	**P**	*0.578*	*0.676*	*0.510*	*0.692*
	R	*0.564*	*0.680*	*0.521*	*0.663*
	F	*0.571*	*0.678*	*0.515*	*0.677*
	P	*0.686*	*0.670*	*0.534*	*0.670*

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

Feature1 and three combined as Feature5	**R**	*0.677*	*0.670*	*0.525*	*0.712*
	F	*0.681*	*0.671*	*0.529*	*0.690*
Feature 2 and 3 combined as Feature6	**P**	*0.678*	*0.672*	*0.615*	*0.671*
	R	*0.663*	*0.678*	*0.620*	*0.719*
	F	*0.671*	*0.675*	*0.617*	*0.694*
Feature 1,2,3 as Combined	**P**	*0.496*	*0.551*	*0.447*	**0.764**
	R	*0.718*	*0.708*	*0.487*	**0.758**
	F	*0.587*	*0.620*	*0.466*	**0.760**

Using feature1, The Maximum of 0.768 as Recall is obtained, which is just a tiny increase of 0.40 over the NB model, but the overall F-measure is 0.760 for RF compared to SVM, obtaining 0.620. According to these findings, multi-base models using a maximum likelihood meta-classifier present a viable method for detecting and communicating multi-class suicidal thoughts in short, "noisy," informal communications like tweets. The fact that "none of the above" is so ambiguous implies that not all hidden problems may be covered by our list of class labels. Categorizing these can be helpful and relevant for future research. Table 7 displays P, R, and F for the best classification model.

Table 6: The optimal classification model's confusion matrix

Class	C1	C2	C3	C4	C5	C6	C7
C1	**71**	0	17	0	0	0	6
C2	2	**26**	2	5	3	5	1
C3	12	0	**151**	0	2	6	18
C4	1	5	6	**21**	2	3	7
C5	2	2	2	0	**38**	2	0
C6	0	5	8	7	2	**82**	5
C7	19	2	21	2	1	3	**62**

7. Discussion

The components are analyzed by applying PCA for all combinations of features; the Union of all three features (Combined set) have produced better results, as shown in Table 4 to Table 7. PCA has reduced 1582 features to 312 features in all terms.

Table 7: Best classification model: P, R, and F

Class	P	R	F
C1	0.731	0.755	0.743
C2	0.650	0.591	0.620
C3	0.729	0.799	0.763
C4	0.601	0.467	0.526
C5	0.808	0.826	**0.817**
C6	0.811	0.752	0.780
C7	0.620	0.564	0.591

8. Conclusion

In order to detect text displaying suicidal intent on Twitter, this study presents several categorization algorithms. Postings about suicidal ideation and topics related to suicide reporting support, advocacy,

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

and memorials are flagged by the model as disturbing. Structural, lexical, psychological, and emotional elements are taken out of Twitter messages to build a base classifier. This classifier was improved by creating an ensemble classifier using the proposed RF method and attained an overall P-0.764, R-0.758 and F of 0.760 for suicidal ideation and Recall-0.830, F-measure- 0.821 and Precision-0.813, for all classes. In order to summarize the results of the study, each class's most crucial predictive PC is highlighted to provide awareness in the language of suicidal communication on Twitter. From this proposed work, it is found that regex and words derived from online discussion forums related to suicide and other micro-blogging websites seem capable of recording particular "clues" in language, such as n-grams, single phrases, and complicated patterns. This method requires additional progress in development with social media text and applications other than Twitter on all platforms. Throughout the investigation and interpretation stages of the study, researchers should maintain close collaboration with experts in the field of "suicidology" to leverage prior understanding of suicidal language features to increase classification accuracy in the future.

References

[1] J. Pirkis & R. W. Blood. (2001). Suicide and the media. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 22(4):155-162.
[2] M. Gould, P. Jamieson, & D. Romer. (2003). Media contagion and suicide among the young. American Behavioral Scientist, 46(9):1269–1284.
[3] K. Daine, K. Hawton, V. Singaravelu, A. Stewart, S. Simkin, & P. Montgomery. (2013). The power of the web: A systematic review of studies of the influence of the internet on self-harm and suicide in young people. PloS one, 8(10): e77555.
[4] D. Baker & S. Fortune. (2008). Understanding self-harm and suicide websites. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 29(3):118–122.
[5] P. R. Recupero, S. E. Harms, & J. M. Noble.Googling. (2008). suicide: surfing for suicide information on the internet. Journal of Clinical Psychiatry.
[6] N. Jacob, J. Scourfield, & R. Evans. (2014). Suicide prevention via the internet: A descriptive review. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 35(4):261.
[7] T. D. Ruder, G. M. Hatch, G. Ampanozi, M. J. Thali, & N. Fischer. (2011). Suicide announcement on Facebook. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 32(5):280-282.
[8] Ikunaga, S. R. Nath, & K. A. Skinner. (2013). Internet suicide in Japan: A qualitative content analysis of a suicide bulletin board. Transcultural psychiatry, page 1363461513487308.
[9] P. Matykiewicz, W. Duch, & J. Pestian. (2009). Clustering semantic spaces of suicide notes and newsgroup articles. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, Association for Computational Linguistics, 179-184.
[10] K. Becker & M. H. Schmidt. (2005). When kids seek help online: Internet chat rooms and Suicide. Reclaiming children and youth, 13(4):229-230.
[11] M. T. Lehrman, C. O. Alm, & R. A. Proano. (2012). Detecting distressed and non-Distressed affect states in short forum texts. In Proceedings of the Second Workshop on Language Social Media, Association for Computational Linguistics, 9-18.
[12] H.-H. Won, W. Myung, G.-Y. Song, W.-H. Lee, J.-W. Kim, B. J. Carroll, & D. K. Kim. (2013). Predicting national suicide numbers with social media data. PloS one, 8(4):e61809.
[13] J. F. Gunn & D. Lester. (2012). Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death. Present tense, 27(16):42.
[14] M. De Choudhury, S. Counts, E. J. Horvitz, & A. Ho. Characterizing. (2014). predicting Postpartum 83 depression from shared Facebook data. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, 626-638.
[15] Merchant RM, Asch DA, Crutchley P, Ungar LH, Guntuku SC, Eichstaedt JC, et al. Evaluating the predictability of medical conditions from social media posts. PLoS ONE 14(6): e0215476, 2019. https://doi.org/10.1371/journal.pone.0215476.
[16] A. Pak & P. Paroubek. (2010) Twitter as a corpus for sentiment analysis and opinion Mining. International Conference on Language Resources and Evaluation.
[17] J. Pennebaker, M. Francis, & R. Booth. (2001). Linguistic Inquiry and Word Count: A We have a computerized text analysis program.
[18] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, & A. Leenaars. (2010). Suicide note Classification using natural language processing: A content analysis. Biomedical informatics insights, (3):19, 2010.

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

[19] J. P. Pestian, P. Matykiewicz, M. Linn-Gust, B. South, O. Uzuner, J. Wiebe, K. B. Cohen, J. Hurdle, & C. Brew. (2012). Sentiment analysis of suicide notes: A shared task. Biomedical informatics insights, 5(Suppl 1):3.

[20] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, & T. McAllister. (2014). Predicting the risk of suicide by analyzing the text of Clinical notes. PloS one, 9(1):e85733.

[21] J. Jashinsky , S.H. Burton , C.L. Hanson , J. West , C. Giraud-Carrier , M.D. Barnes & T.Argyle. (2013). Tracking suicide risk factors through Twitter in the US., Crisis J 35, 51–59.

[22] Sano Y, Takayasu H, Havlin S, Takayasu M, (2019). Identifying long-term periodic cycles and memories of collective emotion in online social media. PLoS ONE 14(3): e0213843. https://doi.org/10.1371/journal.pone.0213843

[23] R. Kavuluru , M. Ramos-Morales , T. Holaday , A.G. Williams , L. Haye , & J. Cerel. (2016). Classification of helpful comments on online suicide watch forums., in BCB, 32–40.

[24] G.B. Colombo, P. Burnap, A. Hodorog, & J. Scourfield. (2016). Analyzing the connectivity and communication of suicidal users on Twitter, Computer Communication, Online Social Networks, 73, 291–300. doi: 10.1016/j.comcom.2015.07.018.

[25] M. Kumar, M. Dredze, G. Coppersmith, & M. De Choudhury. (2015). Detecting changes in suicide content manifested in social media following celebrity suicides, in Proceedings of the 26th ACM Conference on Hypertext & Social Media, ACM, 85–94.

[26] M. Ueda , K. Mori , T. Matsubayashi , & Y. Sawada. (2017). Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on Twitter and subsequent in-ceases in actual suicides, Soc. Sci. Med., 189:158-166.

[27] J.F. Gunn & D. Lester. (2012). Twitter postings and suicide: an analysis of the postings of a fatal suicide in the 24 hours prior to death, Present Tense, 31: 671–678.

[28] J.W. Pennebaker , M. Francis , & R. Booth. (2001). Linguistic Inquiry and Word Count: A computerized text analysis program. doi:10.1.1.600.7227&rep=rep1&type=pdf.

[29] C. Homan , R. Johar , T. Liu , M. Lytle , V. Silenzio , & C. Ovesdotter Alm. (2014). Toward macro-insights for suicide prevention: analyzing fine-grained distress at scale, in: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology, Association for Computational Linguistics, Baltimore,107–117.

[30] Yazdavar AH, Mahdavinejad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al., Multimodal mental health analysis in social media. PLoS ONE 15(4): e0226248, 2020. https://doi.org/10.1371/journal.pone.0226248

[31] L. Biddle, J. Donovan, K. Hawton, N. Kapur, & D. Gunnell. (2008). Suicide and the Internet. BMJ, 336(7648):800-802.

[32] E. Rajesh Kumar, K.V.S.N. Rama Rao, Soumya Ranjan Nayak & Ramesh Chandra, (2020). Suicidal ideation prediction in twitter data using machine learning techniques, Journal of Interdisciplinary Mathematics, 23:1, 117-125.

[33] S. Hemamalini ,V. D. Ambeth Kumar ,R. Venkatesan,S. Malathi. "Relevance Mapping based CNN model with OSR-FCA Technique for Multi-label DR Classification", Fusion: Practice andApplications, Vol. 11, No. 2, 2023 ,PP. 90-110

[34]C. S. Manigandaa,V. D. Ambeth Kumar,G. Ragunath,R. Venkatesan,N. Senthil Kumar. "De-Noising and Segmentation of Medical Images using Neutrophilic Sets", Fusion: Practice and Applications, Vol.11, No. 2, 2023 ,PP. 111-123.

[35] V. Sathya Preiya,V. D. Ambeth Kumar,R. Vijay,Vijay K.,N. Kirubakaran. "Blockchain-Based E-Voting System with Face Recognition", Fusion: Practice and Applications, Vol.12, No. 1, 2023 ,PP. 53-63

[36] Balakrishnan, C.; Ambeth Kumar, V.D. IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks. Diagnostics 2023, 13, 775. https://doi.org/10.3390/diagnostics13040775.

[37] Noora Hani Sherif, Eay Fahidhil, Najlaa Nsrulaah Faris, Hussein Alaa Diame, Raaid Alubady, Seifedine Kadry, Modeling Sports Event Tasks in Augmentative and Alternative Communication Using Deep Learning, Journal of Intelligent Systems and Internet of Things, Vol. 9 , No. 2 , (2023) : 93-107 (Doi : https://doi.org/10.54216/JISIoT.090207).

[38] Tamarah Alaa Diame, Kadim A. Jabbar, Ahmed Taha, Naseer Ali Hussien, Sura Rahim Alatba, Mohammed Nasser Al-Mhiqani, Venkatesan Rajinikanth, Anomaly Detection in Complex Power Grid using Organic Combination of Various Deep Learning (OC-VDL), Journal of Intelligent Systems and Internet of Things, Vol. 9 , No. 2 , (2023) : 78-92 (Doi : https://doi.org/10.54216/JISIoT.090206)

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023

Doi: https://doi.org/10.54216/FPA.140109
Received: September 20, 2023 Revised: October 22, 2023 Accepted: November 11, 2023