SUICIDAL TENDENCY DETECTION

CH. Deepthi[1], Ms.Ch.Vsnprasanna, [2]

#1 Assistant. Professor in the department of CSE at QIS College of Engineering &Technology (Autonomous), Vengamukkapalem(V), Ongole, Prakasam dist.

#2 PG Scholar, Department of Master of Computer Applications, QIS College of Engineering & Technology, Vengamukkapalem, Ongole, Prakasam dist.

ABSTRACT: Suicidal Tendency or the intension to kill oneself or end one's life is a catastrophic situation which is mostly unknown by any person in the victim's life. Suicide has been an intractable public health problem despite advances in the diagnosis and treatment of major mental disorders. In many studies it is clearly evident that, victims tend to kill themselves either to end their pain or pressure or to have a sense of relief that they are not going to live in this world anymore. This project aims to propose a method that helps the family, friends or the close ones of the victim to immediately detect if the person has already started feeling the sense of depression. The main aim is to find a strong co-relation between components in the subsystem and compare the accuracies to build an alarming system. "Better late than never" the victim can be saved by the proposed method and immediate treatment can be started. Unlike the existing systems, this project aims to
detect the suicidal tendencies in multiple aspects instead of focusing on a single perspective

I.INTRODUCTION

This project aims to focus on the people who have an intension to kill themselves. In this consequence, multi-faceted method that can detect this tendency and intimate the family, friends or the close ones beforehand can prove to be a boon for the invention.

This project tends to consider an electronic device particularly a mobile (as used by most of them) is the key element. This device is used to capture different elements like facial gestures, speech recognition and many more. A trivial concept of incorporating different aspects like: Facial Gestures, Voice Recognition and Messaging Patterns follow the bandwagon along with the technical biproducts of the project. Facial gestures include unhappy expressions like sad, dull, tired; voice patterns include low voices that sound dull are easy to recognize that someone is unhappy; Texting patterns include unusual texting patterns that indicate lack of interest in doing activities.

- The objective of this system is that it is capable of detecting suicidal tendency in a specific person. Unlike other existing system this system can ensure the ability to focus on different technical aspects rather than single one.
- The project aims to find co-relation between the three components present below.
- Facial Gesture Detection – Human Computer Interaction
- Speech Recognition – Natural Language Processing

- Messaging Patterns – Text Tokenization through NL

II. EXISTING SYSTEM

Since suicide is not a decision made up in a single day, many researches had been done in the past on

various aspects. While some of them included how suicides occur and others concluded with some methodologies on how to detect suicidal intension. Distinct approaches using Human Computer Interaction, Natural Language Processing and using Convolutional Neural Networks were highly popular. But these systems possessed a drawback that, it was not practically possible to indicate a clear demarcation in only one aspect of implementation. In course of time, numerous approaches have been proposed with the advancement in the technology which is to perform text mining and sentiment analysis on social media platforms like reddit and twitter.

III. PROPOSED SYSTEM

This project aims to propose a system that is capable of detecting suicidal tendency using multiple approaches. Three technologies namely Human Computer Interaction, Natural Language Processing and voice pattern analysis are taken into consideration to perform a detailed execution. Later a correlation matrix is proposed to be implemented that is able to find strong or weak correlations between the above three components.

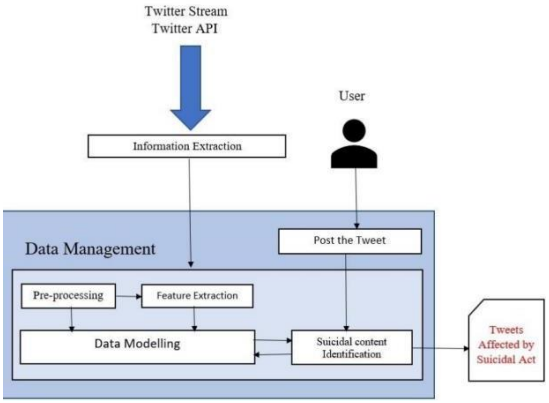
IV. IMPLEMENTATION:

- Data Collection Module: The data collection module is tasked with gathering relevant data related to suicidal tendencies. This data can be sourced from various channels such as surveys, questionnaires, social media platforms, or healthcare records. The module ensures that the collected data is comprehensive and representative of the target population, adhering to ethical guidelines and privacy regulations.
- Data Preprocessing Module: The data preprocessing module is responsible for preparing the collected data for analysis and modeling. It encompasses several tasks, including data cleaning, feature engineering, normalization or scaling of features, and encoding categorical variables into numerical representations. Additionally, it addresses challenges such as handling missing values, outliers, and imbalanced class distributions to ensure the quality and integrity of the dataset.
- Data Splitting Module: The data splitting module divides the preprocessed data into separate subsets for training and testing purposes. Typically, a certain proportion of the dataset is allocated for training machine learning models, while the remainder is reserved for evaluating model performance. By splitting the data, the module enables the assessment of model generalization and effectiveness on unseen data, facilitating robust model development.
- Machine Learning Module: The machine learning module applies advanced algorithms to the training data to construct predictive models for detecting suicidal tendencies. This module incorporates

techniques such as Convolutional Neural Networks (CNNs) and Random Forests. For instance, in CNN implementation, deep learning frameworks like TensorFlow or Py Torch are utilized to design and train models capable of processing diverse data types such as textual or image- based inputs. Conversely, Random Forest implementation involves leveraging machine learning libraries like scikit-learn to create and optimize classifiers based on decision trees, utilizing preprocessed features for training.

- Prediction Module: The prediction module leverages the trained machine learning models to make predictions on new or unseen data instances. Once the CNN and Random Forest models are trained using the respective machine learning module, they are deployed to predict suicidal tendencies based on features extracted from new data points. This module serves as the final stage in the pipeline, enabling the application of predictive models in real-world scenarios to identify individuals at risk of suicidal behavior and facilitate timely intervention and support.

The method proposed in this project first identifies the suicidal keywords, and then we use these keywords to extract tweets from Twitter using Twitter Streaming Application Programming Interface (API).After that we pre-process the data or extract features from the text document.Then SVM and Decision Tree with three types of weight optimizers are imposed on the dataset and at last we determine the model efficiency on the basis of accuracy score, precision, recall and F1 scores.Tweets are retrieved from twitter data and intensity of suicidal tendency is calculated based on the weightage given to the word. Among the two types of machine learning approach we followed supervised learning for this proposed method. A flowchart that represents the proposed model is as follows



Results of classification results using features based on tweets. When we use only features based on tweets, below table shows the accuracy, precision and recall, and Completion Speed of our model in recognizing suicidal posts on tweeter. A 10-fold cross validation was used in the experiments. Table shows that the Decision Tree classifier achieves the best accuracy of 97.89%, precision of 97.89%, recall 97.38% and Naïve Bayes classifier complete a speed of 5.35 second which is the fastest speed among the remaining algorithms.

Algorithms Implemented	Accuracy	precision	recall	Completion Speed (unit in second)
Support vector Machine	92.89%	93.53%	91.41%	1105.52
Decision Tree	97.89%	97.89%	97.38%	18.70
Naïve Bayes	93.29%	89.23%	89.04%	5.35

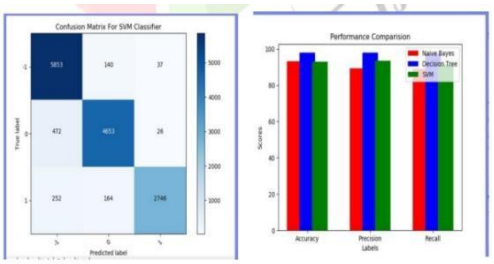


Fig.1. Output results.

Peoples are taking lots of stress due to today's competitive environment in almost all fields like working area and educational area diverting peoples into stress which sometime causes them to take suicide steps. To overcome from this problem author of this paper using PATIENTS dataset to train machine learning algorithms and then this trained model can take patient current status values as input and then predict whether that patient is having any suicidal thoughts. If suicidal thoughts detected then doctors will take necessary steps to help patients in recovering from those thoughts.

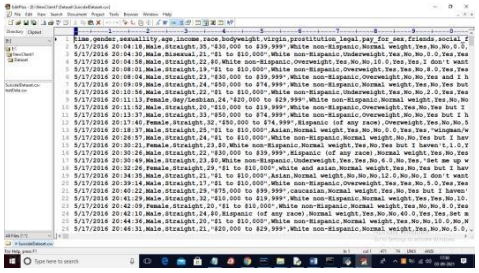
To implement this project author has used KAGGLE suicidal attempt and stress dataset and then train this dataset with deep learning algorithm

called CNN. This dataset contains NON-NUMERIC characters which will not understand by MACHINE LEARNING algorithms so author applying Neural Machine Translation algorithm called Natural Language Processing (NLP) to translate all non-numeric characters to numeric characters which will understand by machine learning algorithms.

In this paper we have trained existing Random Forest and propose CNN algorithm to train dataset and then evaluate both algorithms performance in terms of accuracy, precision, recall and FSCORE.

Note: u r asking to record voice and facial expression to detect depression or suicidal tendency but we don't any senor or devices to record so things so we are using depression and suicidal dataset from KAGGLE which contains columns to detect depression. From facial expression suicide depression can be detected as patient can be depress for some other reason not for suicide so we need to used accurate suicide or depression related dataset to detect such tendency so we choose below dataset

Below is the dataset screen used in this project



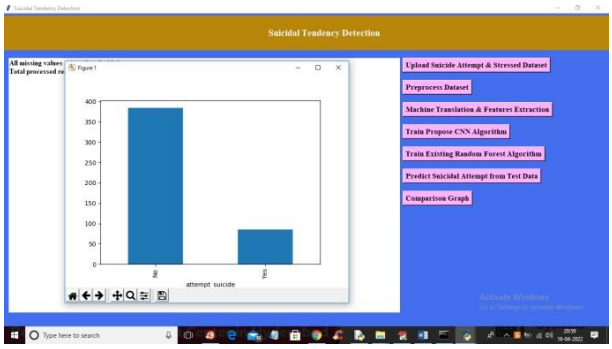
first row contains dataset column names and remaining rows contains dataset values and in above dataset we can see some values are numeric and some are non-numeric and this non-numeric characters will be translate to numeric format by using NLP technique. NLP will assign numeric ID to each unique non-numeric characters and this ID's will be used to train ML algorithms. In below

screen we can see dataset contains 'suicidal attempt' and 'stressed' column.

Then we can see dataset contains 'attempt suicide' and 'stressed' column. We will used above dataset to train ML algorithms. To run project double click on 'run.bat' file to get below screen

Then 'Upload Suicide Attempt & Stressed Dataset' button to upload dataset and to get below screen selecting and uploading 'Suicidal' dataset and then click on 'Open' button to load dataset

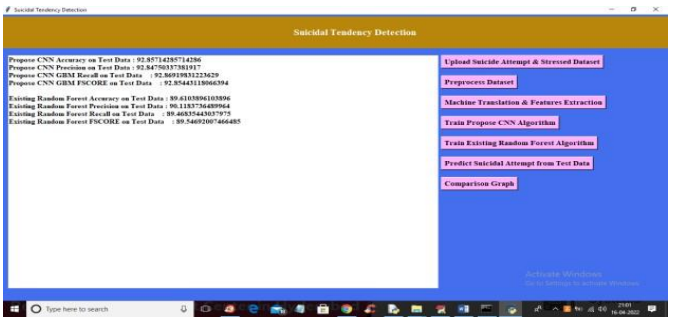
After that we can see some records from dataset and dataset contains some non-numeric characters and to translate them first click on 'Preprocess Dataset' button to remove missing values and then replace with 0



In above screen we can see all missing data is replaced with 0 and we can see dataset contains total 469 records. In graph we can see total patients with and without suicidal thought. In above graph X-axis represents YES and NO values and y-axis represents total counts of YES and NO patients. YES means patients has suicidal thoughts and NO means patients has no suicidal thoughts. Now close above graph and then click on 'Machine Translation & Features Extraction' button to translate all dataset NON-NUMERIC features to NUMERIC features.

In above screen we can see complete dataset is translated to numeric data and here two lines we can see dataset using 614 records to train CNN algorithms and using 154 records for testing CNN

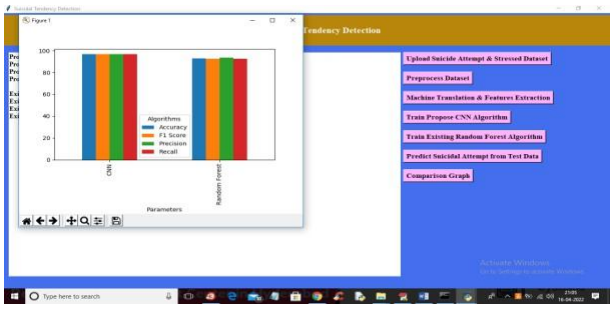
performance. Now train and test data is ready and now click on 'Train Propose CNN Algorithm' we can see with CNN we got 92% accuracy and now click on 'Train Existing Random Forest Algorithm' button to train existing Random Forest algorithm on same data and calculate accuracy



In above screen with existing random forest algorithm we got 89% accuracy and now click on 'Predict Suicidal Attempt from Test Data' button to upload test data and then CNN will predict whether test patient records has any suicidal and NO suicidal thoughts.

We have to select and uploading 'testData.csv' file and then click on 'Open' button to load test data and to get prediction result.

In square brackets we can see patient test records values and after arrow symbol '====>' we can see prediction result from CNN as whether patient has suicidal thoughts or NOT. Now click on 'Comparison Graph' button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy, precision, recall and FSCORE in different color bars. In above graph we can see CNN is performing well compare to existing Random Forest algorithm.

V. ALGORITHMS:

1. Convolutional Neural Network (CNN):

A Convolutional Neural Network (CNN) is a specialized type of deep learning model designed for processing and analyzing visual data, such as images and videos. CNNs are particularly effective due to their ability to automatically learn hierarchical representations of features directly from raw data. The network architecture of a CNN typically consists of convolutional layers, which apply convolution operations using filters to extract features like edges and textures. These features are then passed through pooling layers to down sample and retain important information. Nonlinear activation functions like ReLU introduce nonlinearity into the network, enabling it to learn complex relationships in the data. CNNs are widely used in tasks such as image classification, object detection, semantic segmentation, and face recognition, where they excel in extracting meaningful patterns and making accurate predictions based on visual input.

```
PSEUDOCODE OF CONVOLUTIONAL LAYER
1 for (l = 0;l<L ;l ++){
2     for (m = 0;m<M ;m ++){
3       for (n = 0;n<N ;n ++){
4           sum = bias[l];
5           for (k = 0;k<K ;k ++){
6                 for (s1 = 0;s1 <S1;s1 ++){
7                       for (s2 = 0;s2 <S2;s2 ++){
8                 sum+=weight[k][l][s1][s2] × input [k][m + s1][n + s2];
9           }}}
10      output [l][m][n] = activation_func(sum);
11 }}}
```

2. Random Forest Algorithm:

The Random Forest algorithm is an ensemble learning method used for both regression and classification tasks. It operates by constructing multiple decision trees during training, where each tree is trained on a random subset of the dataset and

a random subset of features. This ensures that each tree has a slightly different view of the data, promoting diversity and reducing overfitting. During prediction, Random Forest aggregates the predictions of all trees to output the mode (for classification) or the average (for regression) of individual tree predictions. This ensemble approach improves predictive accuracy and robustness compared to a single decision tree. Random Forests are versatile and effective across various types of datasets, capable of handling both numerical and categorical data. They are commonly used for tasks such as classification of spam emails, prediction of housing prices, and identifying important features in a dataset through feature importance measures. The algorithm's simplicity, scalability, and ability to provide insights into feature importance make it a popular choice in machine learning applications where

interpretability and performance are critical.

Algorithm 1: Pseudo code for the random forest algorithm

To generate c classifiers:
for i = 1 to c **do**
  Randomly sample the training data D with replacement to produce D_i
  Create a root node, N_i containing D_i
  Call BuildTree(N_i)
end for

BuildTree(N):
if N contains instances of only one class **then**
  return
else
  Randomly select x% of the possible splitting features in N
  Select the feature F with the highest information gain to split on
  Create f child nodes of N, N_1,..., N_f, where F has f possible values (F_1, ... , F_f)
  for i = 1 to f **do**
    Set the contents of N_i to D_i, where D_i is all instances in N that match
    F_i
    Call BuildTree(N_i)
  end for
end if

VI. RESULT

S.NO	Algorithm	Accuracy	Precision	Recall	FSCORE
1	Random Forest	92.85	92.84	92.86	92.85
2	CNN	89.61	90.11	89.46	89.54

VII. CONCLUSION

Machine Learning applied to detect suicide intention and depression amongst individual is effective as traditional approaches are hindered by factors like face-to-face conversation and shyness to express themselves. In the future I would like to delve more into context analysis via the retweet history and given links to external sites. Even though ML seems to be a better method, yet it entails some shortcomings, like eventually having to use human intervention to approve the predictions by the ML model, it's limited to detecting suicidal tendencies and depression and contextual analysis as the data collected doesn't come with prior background explanation. The amount of text keeps increasing with the popularization of social networking services. And suicide detection and prevention remain a crucial task in our modern society. It is therefore essential to develop new methods to detect online texts containing suicidal ideation in the hope that suicide can be prevented. In the project an automatic recognition of suicidal posts is presented using machine learning techniques As a future scope we can implement some more algorithms to improve the efficiency and quality of work and we can implement the same project on videos using image processing techniques.

VIII. REFERANCES

[1]. Hayes, L.M., 2013. Suicide prevention in correctional facilities: Reflections and next steps. International journal of law and psychiatry 36, 188–194

[2]. S. Lee et al., "Detection of a Suicide by Hanging Based on a 3-D Image Analysis," in IEEE Sensors Journal, vol. 14, no. 9, pp. 2934-

2935, Sept.
2014. doli: 10.1109/JSEN.2014.2332070.
[3]. Calderon-Vilca, H. D., Win-Rafael, W. I., & Mandelate, R. (2017), "Simulation of suicide
tendency by using machine learning", 2017 36th International Conference of the Chilean Computer Science,
Society(SCCC).doi:10.1109/sccc.2017.8405128.
[4]. Hu, Z., Hu, Y., Wu, B., & Liu, J. (2017), "Hand Pose Estimation with CNN-RNN", 2017 European Conference on Electrical Engineering and Computer Science (EECS). doi:10.1109/eecs.2017.91.
[5]. Kamel, A., Sheng, B., Yang, P., Li, P., Shen, R., & Feng, D. D. (2018), "Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures", IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp.1 doi:10.1109/tsmc.2018.2850149.
[6]. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Moore, R. (2013)," Real-time human pose recognition in parts from single depth images", Communications of the ACM, Vol.56(1), 116. doi:10.1145/2398356.2398381.
[7]. Hu, L., & Xu, J. (2017)," Body Joints Selection Convolutional Neural Networks for Skeletal Action Recognition", 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). doi:10.1109/ictai.2017.00109.
[8]. Kim, Y., Kim, M., Goo, J., & Kim, H. (2018), "Learning Self-Informed Feature Contribution for Deep Learning-Based Acoustic Modeling." IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.26(11),2204–2214. doi:10.1109/taslp.2018.2858923
[9].Crombec, N., Caron, G., Fumatoria, T., &Mukaigawa, Y. (2018), &putrefiable Planar Object Pose Estimation in Light Fields From Best Sub aperture Camera Pairs.&Quot; IEEE Robotics and Automation Letters, Vol.3(4),pp.3561–3568. doi:10.1109/lra.2018.2853267.
[10]. Jianhong Wang,1 Tian Lan,, "Spartoi-temporal Aware Non-negative Component Representation for Action Recognition ", IEEE Transactions On Parallel And Distributed Systems, Vol. 26, no. 9, pp. 2520-2533, 2015.

AUTHORS

[1] Mrs. Chepuri Deepthi, currently working as an Assistant professor in the department of Computer Science and Engineering, QIS college of Engineering and Technology, Ongole, Andhra Pradesh. She did her B. Tech from Uttar Pradesh Technical University, Lucknow and M. Tech from JNTUK, Kakinada. Her area of interests is Machine Learning, Artificial Intelligence, Cloud Computing and Programming languages.

[2] Ms. Ch. Vsnprasanna, currently pursuing Master of Computer Applications at QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh. He completed BSC in Computer Science from Chaitanya Bharathi Degree College, Chirala, Ongole, Andhra Pradesh. His areas of interests are Machine Learning.