

Suicidal Tendency Detection Using Machine Learning

B. Channarayapriya¹, P. Suresh kumar Reddy²

¹M.Tech Student, Department of CSE, Golden Valley Integrated Campus, Madanapalli

²Associate Professor, Department of CSE, Golden Valley Integrated Campus, Madanapalli

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 correlation 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.

Keywords: Suicide, Health problem, Depression, Earlier Detection

INTRODUCTION

Suicide is a serious issue in modern society. There are several factors that can lead a person to commit a suicide, for instance, stress, depression, failure, disappointment, pessimism, unemployment, among others. A recent study by the Center for Disease Control (CDC) of the United States of America estimated that the economic toll of suicide on society is immense as well. Suicides and attempts to suicide cost the nation approximately \$70 billion per year in lifetime medical and work-loss costs alone.

Detecting and preventing suicide attempts therefore becomes crucial for the authorities of the land and should be addressed in order to save people's lives and preserve the social fabric of the community. Nowadays social media has become a way for people to express themselves and thus may be used to convey suicidal tendencies. Due to the addiction of social networks, people express their feelings and sufferings in various online communities, modes of prevention include interaction between likely suicidal individuals and experts, therapists or social workers. But this mode of operation is susceptible to create some delays in the diagnosis of the patient, which could lead to a fatality.

In order to prevent suicides more effectively, the ideation must be detected as early as possible. This can be done by analyzing users' posts for suicidal related content. The key objective being to present an automatic recognition of suicidal posts using Machine learning techniques. We focus on the online Twitter users. The project will consist of developing a model of classification of various social media posts into classes that determine whether the user has suicidal tendencies or not. The goal of text-based suicide classification is to determine whether candidates have suicide ideations or not through their posts.

Traditional modes of prevention rely on clinical methods, including self-reports and face-to-face interviews that's interactions between likely suicidal individuals and an expert, a therapist, or a

social worker. The main objectives of this project are an earlier detection of suicidal ideation, an automatic recognition of suicidal content in social networking posts like twitter, reddit.

Suicide is quite possibly the most recognized reasons for death on the news around the world. There are a few elements and factors that can lead an individual to submit this suicide, for instance, stress, confidence, sadness, among others. We propose a new methodology with a machine learning using twitter dataset; such information mirrors the young adult populace with self-destructive inclination in many. We at last propose a technique that decides the suicidal detection in posts.

Self-destruction has consistently had its situation among the best 10 reasons for death everywhere in the world. It was assessed by the World Health Organization (WHO) that consistently roughly 1,000,000 individuals ended it all, which brings the death rate to 16 individuals for each 100,000 or one demise each 40 seconds. It is anticipated that the self-destruction rate is going to ascend to one at regular intervals. More youthful ages have now supplanted older guys as the gathering at most elevated danger in many nations. Psychological well-being messes, (for example, discouragement and substance misuse) are credited with more than 90% of all instances of suicide. Numerous nations perceive the need and positive effect of Suicide Prevention Strategies and are attempting to guarantee they are set up. One such anticipation system is early detection of suicidal ideation or melancholy among people.

The scope of this project is an earlier Detection Of Suicidal Ideation (DSI). The key objective is to present an automatic recognition of suicidal posts using Machine learning techniques. The goal of text-based suicide classification is to determine whether candidates, through their posts, have suicidal ideations. Machine learning methods and NLP also will be applied in this field.

We have used Twitter as our dataset for this project. Twitter provided us with four keys - consumer_key, consumer_secret, access_token, access_secret. The dataset was split into 80% training and 20% testing. The data was provided in text format and we have performed preprocessing to remove unwanted data like emojis, signs (!/@\$:*&^%) and stopwords and then we have stored it in csv format.

LITERATURE SURVEY

1) Suicidal Tendency Detection:

This paper was written by the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain and it was published by IEEE in the year 2019.

They have used the Convolutional Neural Network (CNN) in this project to get an accuracy of 77%. The drawback of this approach was that they have used image based predictive models For detecting depression and suicide.

2) Performance Evaluation of Different Machine Learning Techniques using Twitter Data for Identification of Suicidal Intent:

This paper was written by Anirudh Ramachandran, Akshara Gadwe, Dishank Poddar, Saurabh Satavalekar and Sunita Sahu and was published by IEEE in 2020 at the International Conference on Electronics and Sustainable Communication Systems (ICESC). Research and Evaluation based on online behavior have been conducted repeatedly. Using machine learning, this online trail of data that a person leaves behind can be used to gain insights on the behavior and psychological status. In this paper, different machine learning techniques have been used, studied and gauged their effectiveness for suicidal tendency detection to prove that Machine Learning Algorithms like Logistic Regression can correctly identify residing Suicidal Tendency of a Twitter user. They have used algorithms such as Linear Regression, Logistic Regression, Naive Bayes , Random Forest, GBDT, XGBoost, MLFFNN on the twitter dataset to achieve an accuracy of 76.3%. The only drawback of this approach is that it uses logistic regression which is a long process for detection of suicides

3) DEPRESSION AND SUICIDAL TENDENCY IDENTIFICATION:

This paper was written by Seung Young Ryu ,Hyeongrae Lee,Dong-Kyun Lee AND Kyeongwoon Park Department of Mental Health Research ,National Center for Mental Health, Seoul, Republic of Korea and was published on 30th December 2019.

This approach uses the Random Forest algorithm to get an accuracy of 85%.

4) Simulation of Suicide Tendency by Using Machine Learning:

This paper was written by Hugo D. Calderon-Vilca, William I. Wun-Rafael and Roberto Miranda-Loarte and was published by IEEE on 16th October 2017 at the 36th International Conference of the Chilean Computer Science Society (SCCC).

They have proposed a simulation with a systematically generated dataset. Such data reflect the adolescent population with suicidal tendencies in Peru. They have evaluated three algorithms of supervised machine learning as a result of the algorithm C4.5 which is based on the trees to classify in a better way the suicidal tendency of adolescents. They finally propose a desktop tool that determines the suicidal tendency level of the adolescent. The algorithms used are C4.5, JRip algorithm and Naive Bayes algorithm and have achieved an accuracy of 90.7%. This approach falls short as it was implemented only on adolescents.

5) Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis:

This paper was written by Diana Ramírez-Cifuentes , Ana Freire , Ricardo Baeza-Yates , Joaquim Puntí , Pilar Medina-Bravo , Diego Alejandro Velazquez, Josep Maria Gonfau, Jordi González and was published in the year 2020 in the Journal of Medical Internet Research.

This paper aimed to describe an approach for the suicide risk assessment of Spanish-speaking users on social media. We aimed to explore behavioral, relational, and multimodal data extracted from multiple social platforms and develop machine learning models to detect users at risk. They characterized users based on their writings, posting patterns, relations with other users, and images posted. They also evaluated statistical and deep learning approaches to handle multimodal data for the detection of users with signs of suicidal ideation. To evaluate the performance of the models, they distinguished 2 control groups:

users who make use of suicide-related vocabulary (focused control group) and generic random users (generic control group).

The algorithms used were random forest, multilayer perceptron, logistic regression, and support vector machines as classifiers which gave an accuracy of 82%. But this is just an observational study. Results can be improved by enhancing the contribution of the textual and relational features

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.

Disadvantages :

- Can predict outcome only from one aspect.
- The output is not reliable

PROPOSED SYSTEM

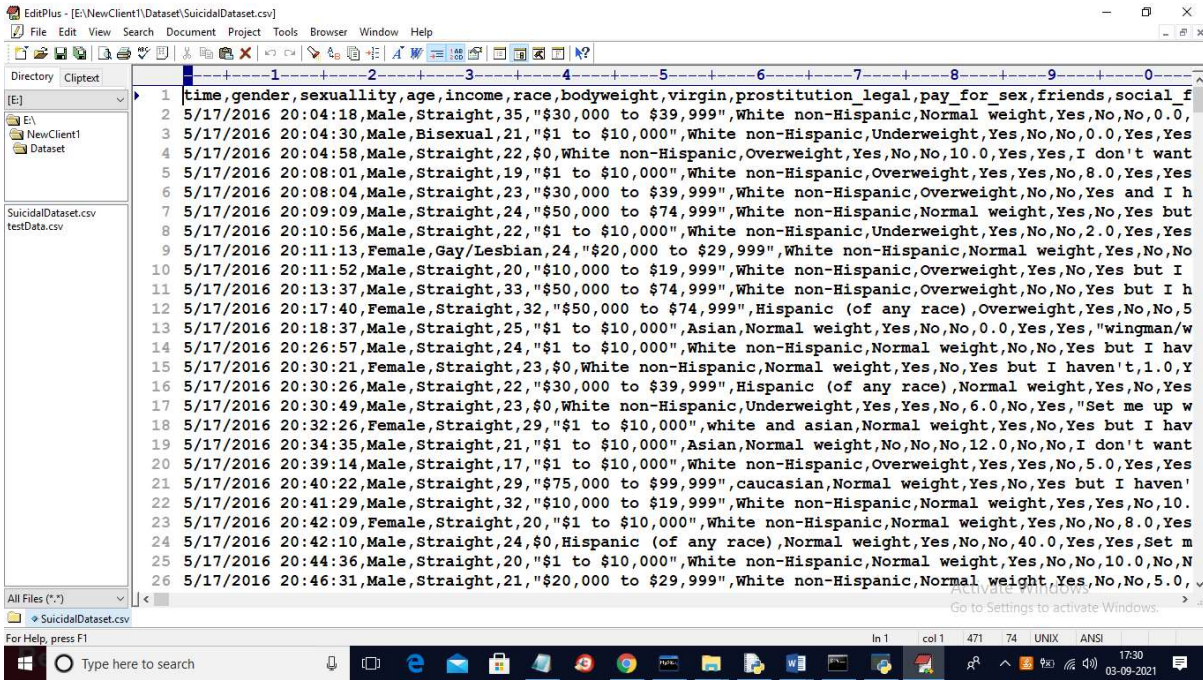
This project aims to propose a system that is capable of detecting suicidal tendency using multiple approaches. Three technologies namely Human Computer Interaction, Natural Language Processing and voice pattern analysis are taken into consideration to perform a

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.

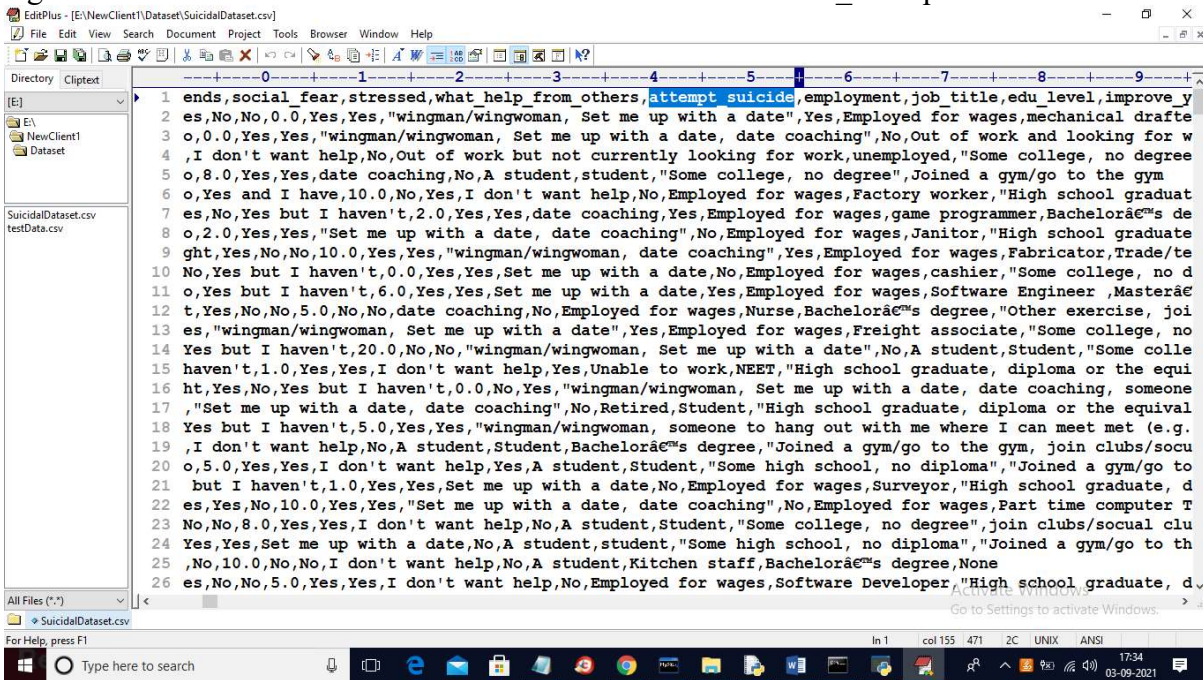
Advantages :

- Predicts outcomes from multiple aspects.
- Reliable outcome due to correlation matrix.

RESULTS

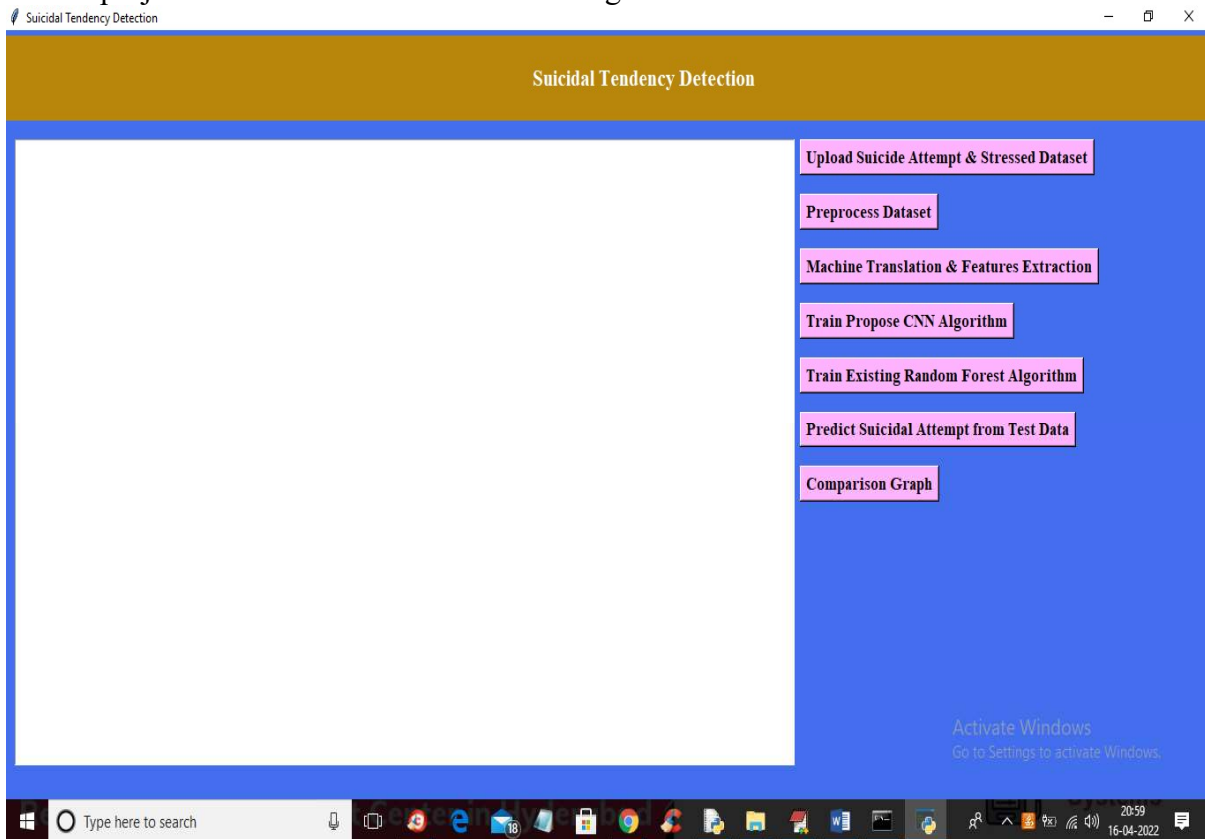


In above dataset first row contains dataset column names and remaining rows contains dataset values and in above dataset we can see some values are numeric and some are non-numeric and this non-numeric characters will be translate to numeric format by using NLP technique. NLP will assign numeric ID to 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.

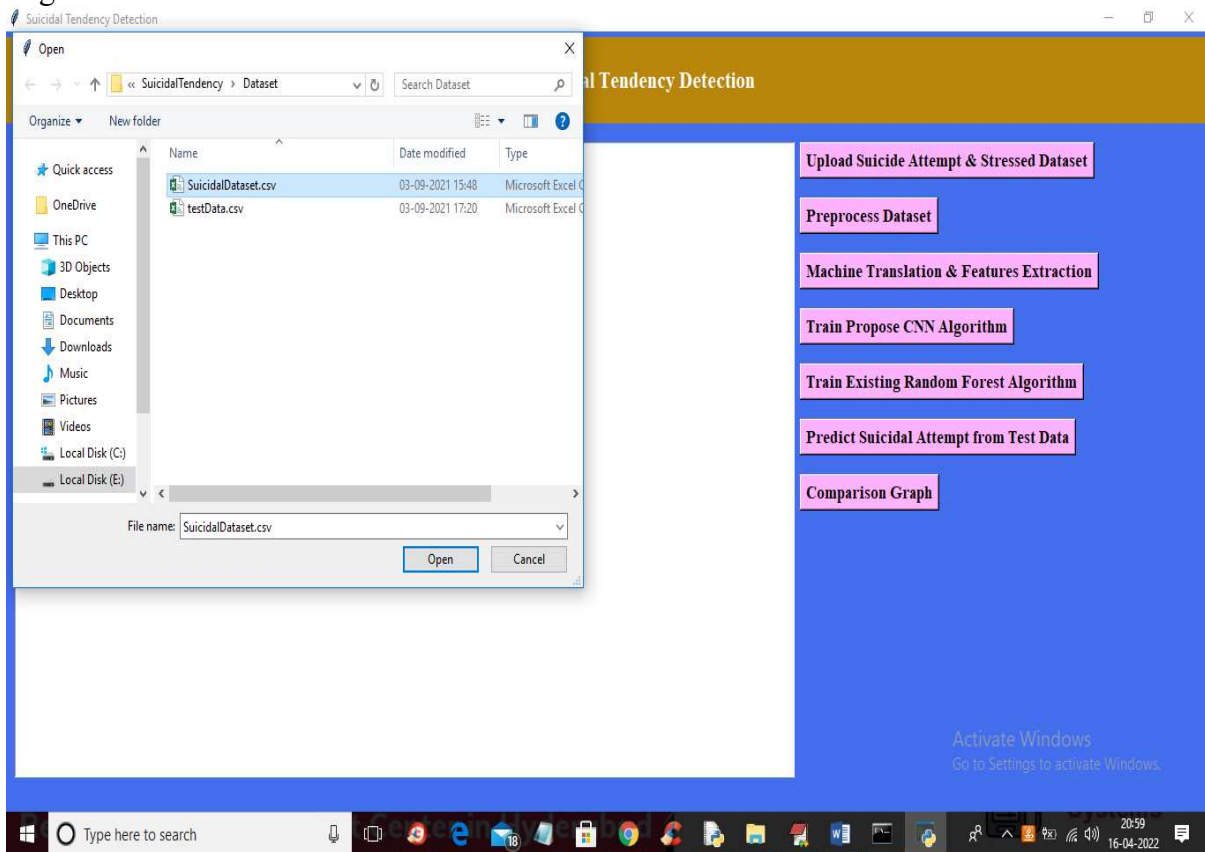


In above dataset screen you can see dataset contains 'attempt_suicide' and 'stressed' column. We will used above dataset to train ML algorithms.

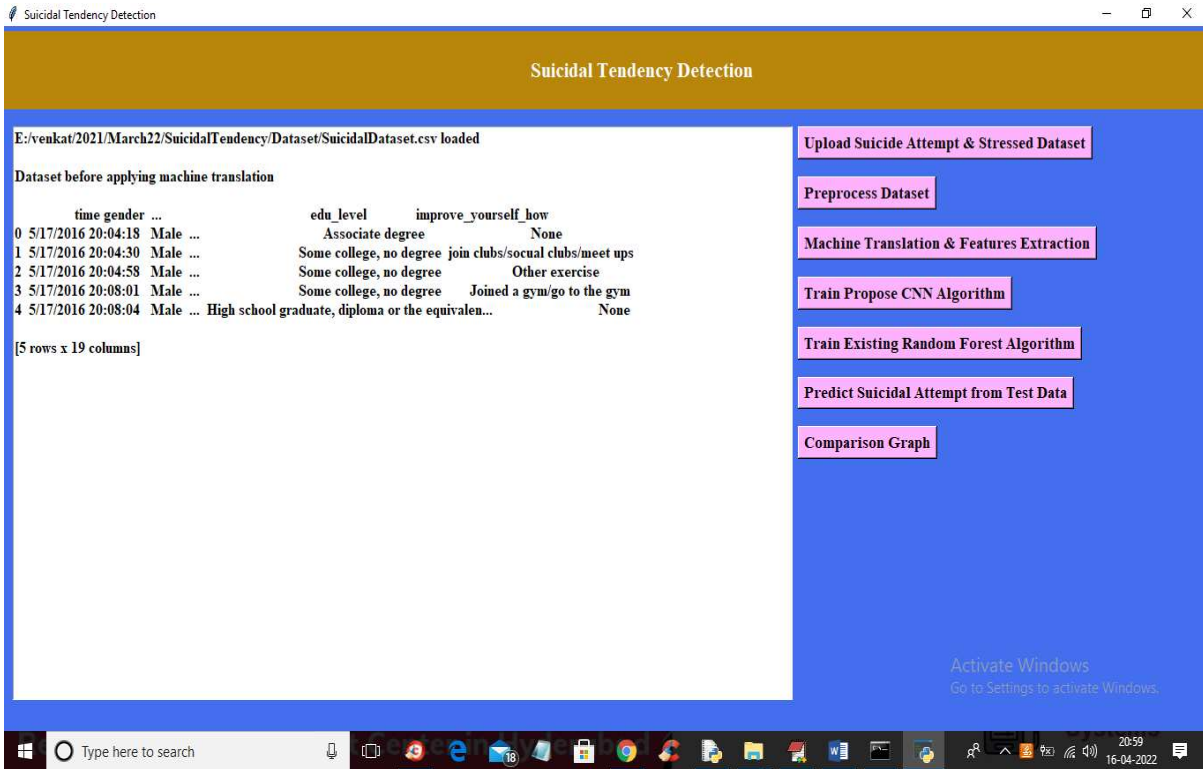
To run project double click on 'run.bat' file to get below screen



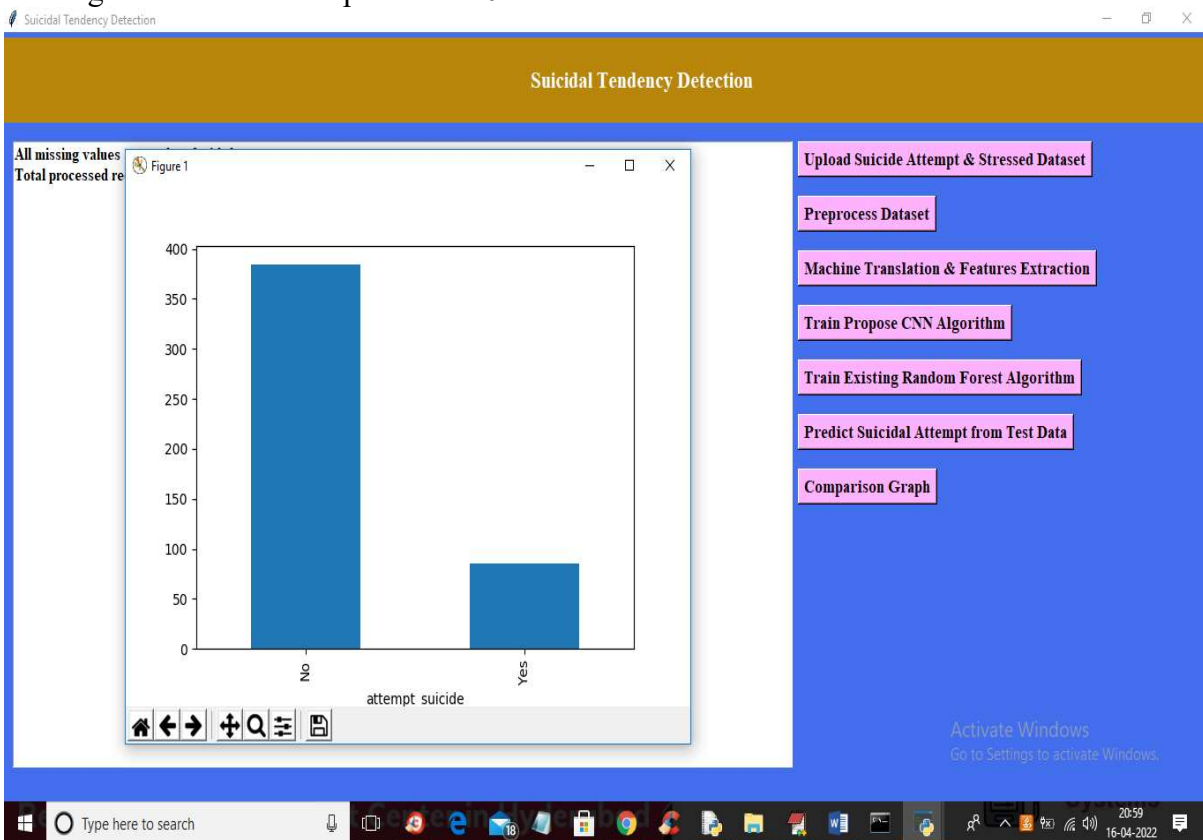
In above screen click on 'Upload Suicide Attempt & Stressed Dataset' button to upload dataset and to get below screen



In above screen selecting and uploading 'Suicidal' dataset and then click on 'Open' button to load dataset and to get below screen

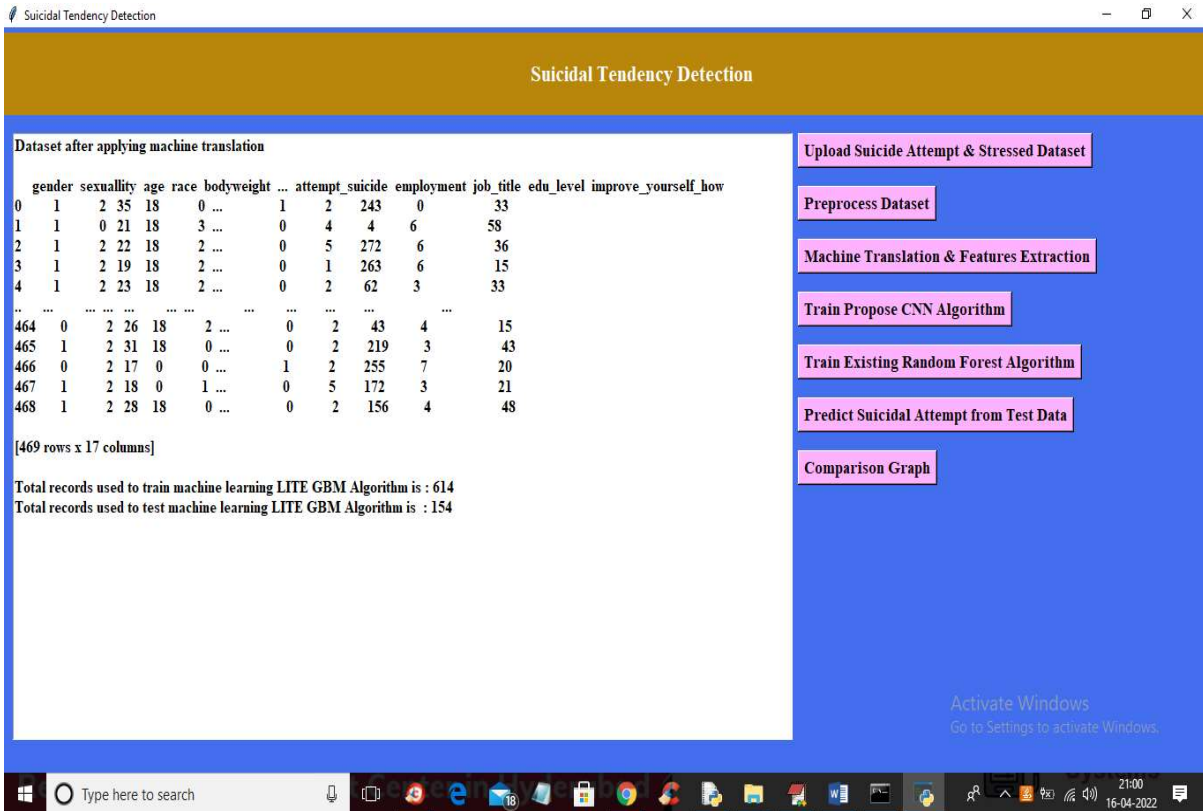


In above screen dataset loaded and we can see some records from dataset and dataset contains some non-numeric characters and to translate them first click on 'Preprocess Dataset' button to remove missing values and then replace with 0



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

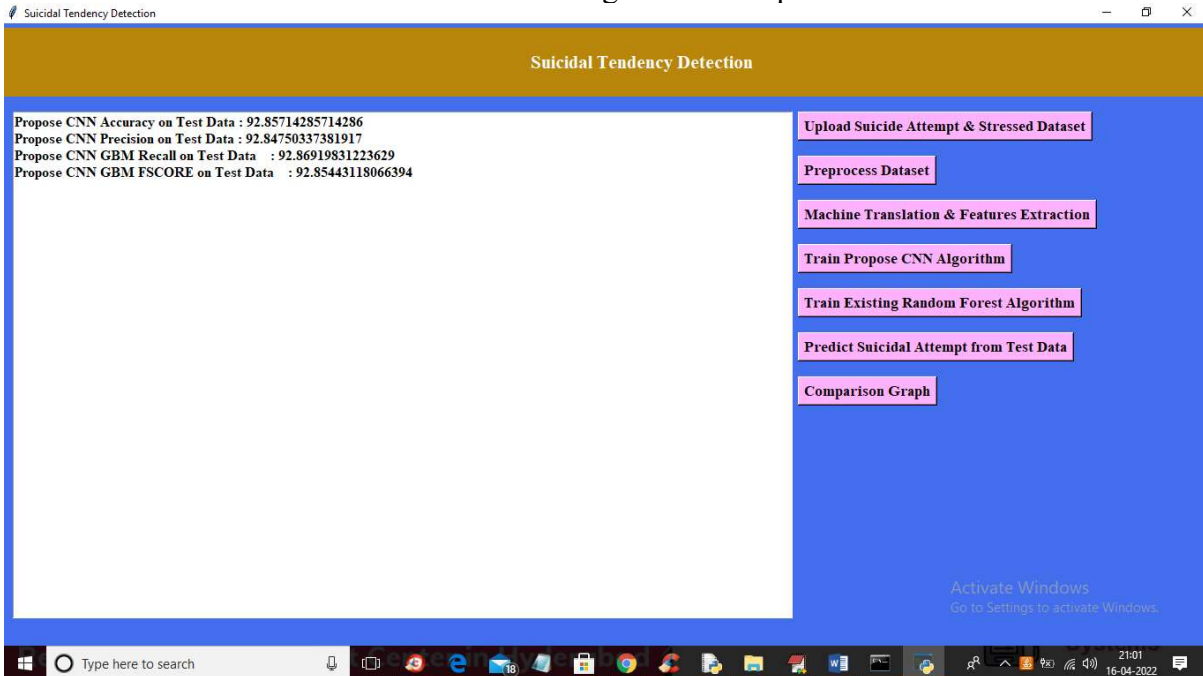
close above graph and then click on ‘Machine Translation & Features Extraction’ button to translate all dataset NON-NUMERIC features to NUMERIC features.



The screenshot shows the 'Suicidal Tendency Detection' application window. The main area displays a table titled 'Dataset after applying machine translation'. The table has 17 columns: gender, sexuality, age, race, bodyweight, ..., attempt, suicide, employment, job_title, edu_level, improve_yourself_how. The first few rows are visible, showing numerical values for each feature. Below the table, it indicates '[469 rows x 17 columns]'. To the right of the table, there are several buttons: 'Upload Suicide Attempt & Stressed Dataset', 'Preprocess Dataset', 'Machine Translation & Features Extraction', 'Train Propose CNN Algorithm', 'Train Existing Random Forest Algorithm', 'Predict Suicidal Attempt from Test Data', and 'Comparison Graph'. At the bottom of the window, there is a search bar and a taskbar with various application icons. The system tray shows the time as 21:00 and the date as 16-04-2022.

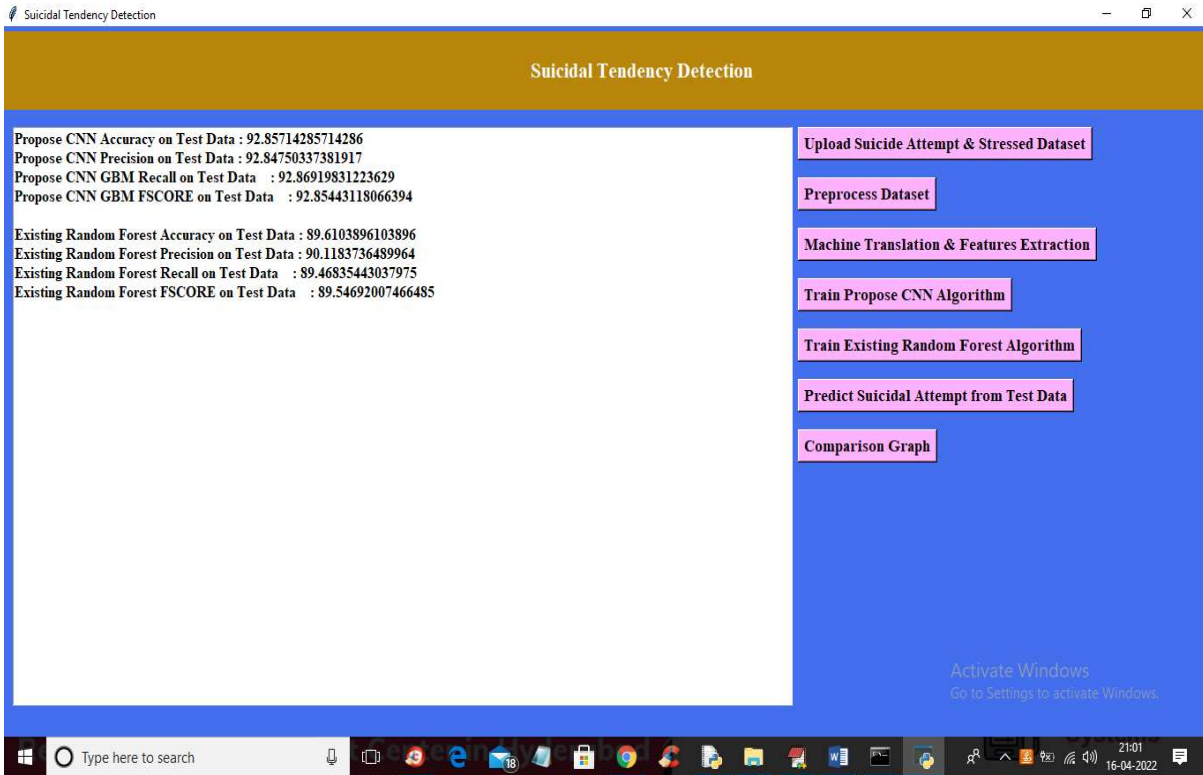
| | gender | sexuality | age | race | bodyweight | ... | attempt | suicide | employment | job_title | edu_level | improve_yourself_how |
|-----|--------|-----------|-----|------|------------|-----|---------|---------|------------|-----------|-----------|----------------------|
| 0 | 1 | 2 | 35 | 18 | 0 | ... | 1 | 2 | 243 | 0 | | 33 |
| 1 | 1 | 0 | 21 | 18 | 3 | ... | 0 | 4 | 4 | 6 | | 58 |
| 2 | 1 | 2 | 22 | 18 | 2 | ... | 0 | 5 | 272 | 6 | | 36 |
| 3 | 1 | 2 | 19 | 18 | 2 | ... | 0 | 1 | 263 | 6 | | 15 |
| 4 | 1 | 2 | 23 | 18 | 2 | ... | 0 | 2 | 62 | 3 | | 33 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 464 | 0 | 2 | 26 | 18 | 2 | ... | 0 | 2 | 43 | 4 | | 15 |
| 465 | 1 | 2 | 31 | 18 | 0 | ... | 0 | 2 | 219 | 3 | | 43 |
| 466 | 0 | 2 | 17 | 0 | 0 | ... | 1 | 2 | 255 | 7 | | 20 |
| 467 | 1 | 2 | 18 | 0 | 1 | ... | 0 | 5 | 172 | 3 | | 21 |
| 468 | 1 | 2 | 28 | 18 | 0 | ... | 0 | 2 | 156 | 4 | | 48 |

In above screen we can see complete dataset is translated to numeric data and in below 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’ button to train CNN with above dataset and to get below output

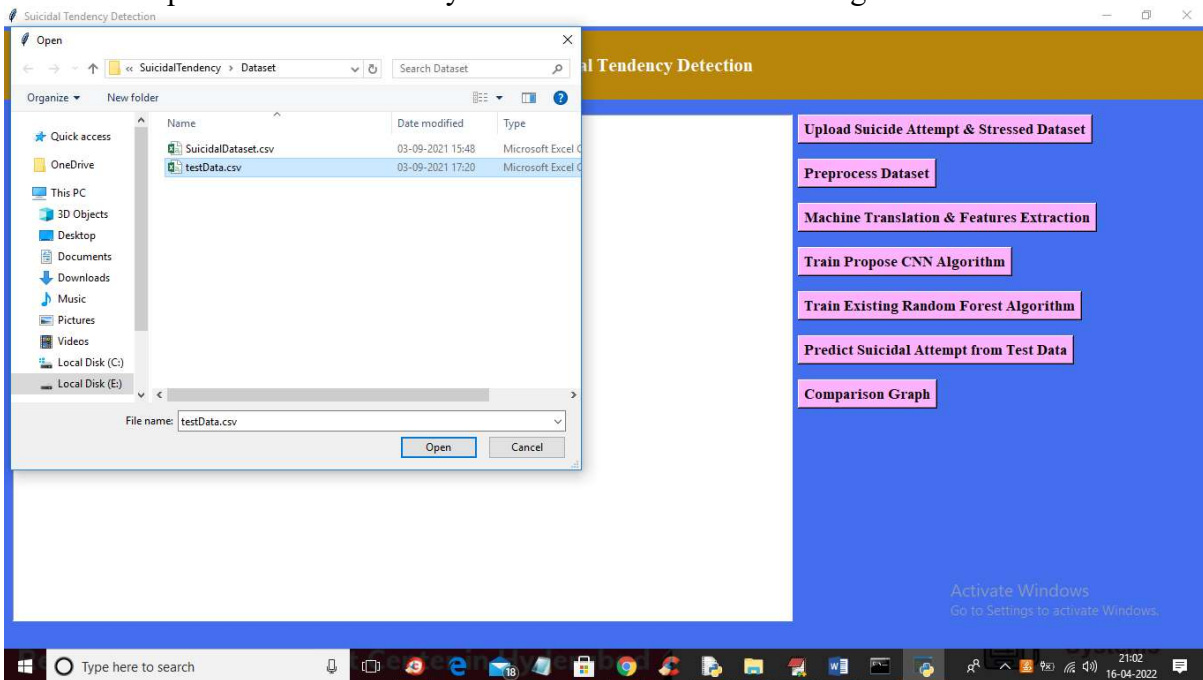


The screenshot shows the 'Suicidal Tendency Detection' application window. The main area displays the output of the CNN algorithm. The text reads: 'Propose CNN Accuracy on Test Data : 92.85714285714286', 'Propose CNN Precision on Test Data : 92.84750337381917', 'Propose CNN GBM Recall on Test Data : 92.86919831223629', and 'Propose CNN GBM FSCORE on Test Data : 92.85443118066394'. To the right of the output, there are several buttons: 'Upload Suicide Attempt & Stressed Dataset', 'Preprocess Dataset', 'Machine Translation & Features Extraction', 'Train Propose CNN Algorithm', 'Train Existing Random Forest Algorithm', 'Predict Suicidal Attempt from Test Data', and 'Comparison Graph'. At the bottom of the window, there is a search bar and a taskbar with various application icons. The system tray shows the time as 21:01 and the date as 16-04-2022.

In above screen 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

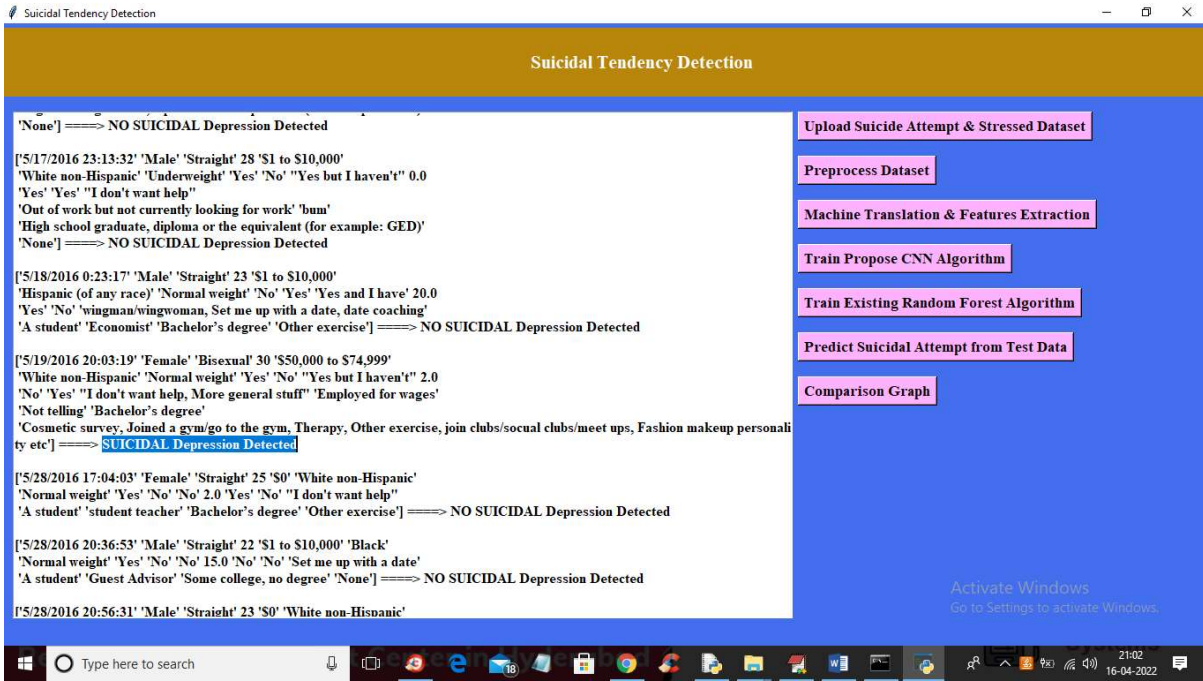


In above screen with existing random forest algorithm we got 89% accuracy and now click on ‘Predict Suicidal Attempt from Test Data’ button to upload test data and then CNN will predict whether test patient records has any suicidal and NO suicidal thoughts.

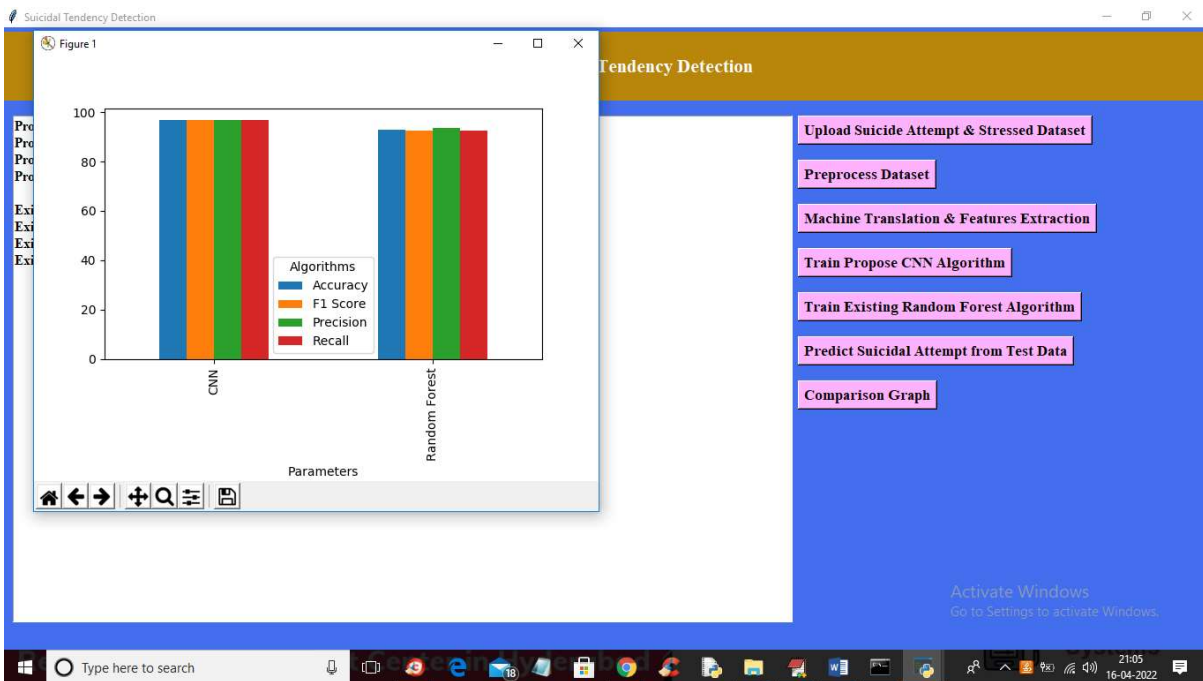


In above screen we are selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to load test data and to get below prediction result

In above screen 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. You can scroll down above screen to get all predicted records like below screen



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 colour bars. In above graph we can see CNN is performing well compare to existing Random Forest algorithm.

CONCLUSION

Deep learning plays a vital role day to day life in hardware, software and various devices. They are used in various applications like virtual personal assistant, traffic predictions, online transportation networks, video surveillance, online fraud detection, refinement of search engine results, product recommendations and social media services.

An innovative technique using machine learning is proposed for detecting hanging attempts. After training the system it recognizes the hanging attempts. It captures the actions through a camera,

generates an alert message. Using this technique better accuracy and higher sensitivity are obtained on a dataset with substantial variations between different simulated hanging sequences.

In future, the focus will be more on enhancing the proposed method to increase the chances of detecting the hanging attempts quickly thereby preventing the suicide attempts. By combining the images captured through surveillance camera and the depth information an intriguing path is used to demonstrate the connection between the hanging person and the strangling object.

REFERENCES:

- [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. doi: 10.1109/JSEN.2014.2332070.
- [3]. Calderon-Vilca, H. D., Wun-Rafael, W. I., & MirandaLoarte, 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.1–14. 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]. Crombez, N., Caron, G., Funatomi, T., & Mukaigawa, Y. (2018), "Reliable Planar Object Pose Estimation in Light Fields From Best Subaperture Camera Pairs." *IEEE Robotics and Automation Letters*, Vol.3(4), pp.3561–3568. doi:10.1109/lra.2018.2853267.
- [10]. Jianhong Wang, Tian Lan,, "Spatio-temporal Aware Non-negative Component Representation for Action Recognition ", *IEEE Transactions On Parallel And Distributed Systems*, Vol. 26, no. 9, pp. 2520-2533, 2015.
- [11]. Lee, M., Shin, S., Hong, S., & Song, S. (2017), "BAIPAS: Distributed Deep Learning Platform with Data Locality and Shuffling.", 2017 European Conference on Electrical Engineering and Computer Science (EECS). doi:10.1109/eecs.2017.10
- [12]. Deb, S., Rabiul Islam, S. M., Johura, F. T., & Huang, X. (2017), "Extraction of Linear and Non-Linear Features of Electrocardiogram Signal and Classification" 2nd International Conference on Electrical & Electronic Engineering (ICEEE). doi:10.1109/ceee.2017.8412857.
- [13]. Wang, Q., Gong, D., Li, M., Zhao, C., & Lei, Y. (2017), "Sparse feature auto-combination deep network for video action recognition", 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). doi:10.1109/fskd.2017.8393360