ISSN: 2348-2079

Volume-11 Issue-1

International Journal of Intellectual Advancements and Research in Engineering Computations

Machine learning based suicide prediction using RNN

Parama Guru. S, B.Narmada

Department of Computer Science and Engineering, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu.

***Corresponding Author: Parama Guru. S**
Email: paramagurucse53@gmail.com, Salem.narmadadgct@gmail.com

ABSTRACT

Machine learning (ML) and artificial intelligence case study supports to increase the accuracy level of prediction and aid the goal of suicide prevention. This paper reviews literature concerning the machine learning methods used to help identify various risk factors and help prevent suicide. This existing method our research and analysis finding wrong prediction which were used to identify various suicide risk factors and additional analysis of whether there are any correlations or variations in the risk factors from pre- and post-pandemic datasets regarding suicide rates low. The proposed method using to human behavior based predict the suicide implementation of Recurrent Neural Network algorithm (RNN) using human behavior data sets and preprocess, classified of the suicide or non-suicide prevention. The dataset obtained from suggest that high levels of risk factor identification are possible and this study and the analysis serve as supporting research and monitor to relief in the continued ambitious goal of suicide prevention. As a result, the focus of this study is to illustrate some of the computational strategies utilized in the framework proposed in this study can increase the accuracy machine learning to predicting at risk of suicide prevention.

Keywords: *Artificial Intelligence (AI), Machine learning (ML), Neural Network (NN), Suicide Non Suicide Prediction.*

INTRODUCTION

Suicide is a serious public health problem with long-lasting effects on individuals, families and communities. The good news is that suicide is preventable. Suicide prevention needs strategies at all levels of society. It includes prevention and protection strategies for individuals, families and communities. By learning the warning signs, promoting prevention and resilience and engaging in social change, everyone can prevent suicide.

Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to mimic how humans learn and to improve their accuracy over time. Machine Learning is a key component of the ever-growing field of Data Science. Use statistical techniques to classify or predict and train algorithms to reveal key insights in data mining projects. These insights then drive decisions in the application and business, ideally influencing key growth metrics. As Big Data expands and continues to grow, the market need for data scientists will increase.

A neural network is an artificial intelligence technique that teaches computers to process data in ways inspired by the human brain. It's a machine learning process called deep learning that uses interconnected nodes, or neurons, in a hierarchy similar to the human brain. It creates an adaptive system that the computer can use to learn from its mistakes and improve continuously. In this way, artificial neural networks try to more accurately solve complex problems, such as document summarization and face recognition.

Neural networks can track user activity and make personalized suggestions. We may also analyze the behavior of all our users and discover new products and services that may be of interest to particular users. For example, Curalate, a Philadelphia-based startup, helps brands convert social media posts into sales. Brands use Curalate's Intelligent Product Tag (IPT) service to automatically collect and curate user-generated social content. IPT uses neural networks to automatically find and recommend products relevant to users' social media activity. Consumers can find specific products from social media images without having to search through online catalogs. Instead, you can use Curalate's automatic product tagging feature to easily shop for products.

One might argue that not all factors in a predictive model contribute equally to determining suicide rates in a particular county. Therefore, there is an urgent need to investigate and identify which variables influenced the predicted outcome and to quantify their contribution37. This is usually achieved using importance scores. The importance score roughly indicates how important each feature is in building the predictive model. The more features a model uses to make a decision, the higher the importance score38. Feature

importance values are computed explicitly for each attribute, allowing attributes to be easily sorted and compared.

METHODOLOGY

In this section we have to discuss the various process of predicting the suicide cases. First we collect dataset and then preprocess the collected dataset, after that we have to extract the features from the dataset and select the best valuable features and finally the features can be classified by using RNN algorithm.

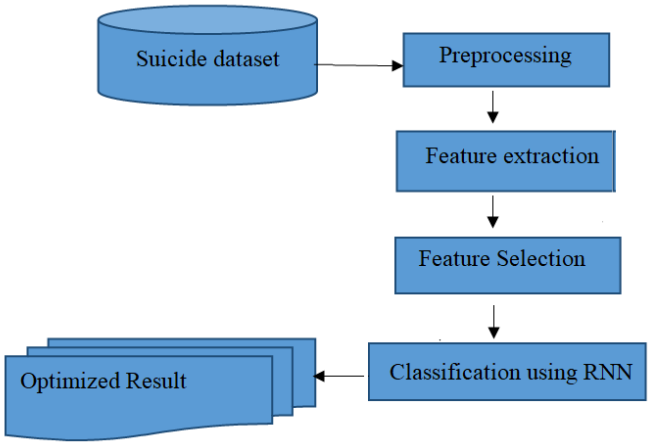

Fig 1.1 Basic flow diagram of suicide prediction

Fig1 described as. Suicide detection is processed in different stages such as pre- Processing, feature extraction, feature selection and classification. Generally, there are four phases to determining whether or not you have suicide. The first phase entails obtaining suicide dataset. In the second phase, data set can be pre-processing to improve the quality. In the third phase, Image extraction can extract the best feature in the trained dataset. Image segmentation is the fourth phase, and finally the data set can be classified and obtained the result.

Dataset collection

The dataset is borrowed from Kaggle, https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016.
This is a compiled dataset pulled from four other datasets linked by time and place from year 1985 to 2016. The source of those datasets is WHO, World Bank, UNDP and a dataset published in Kaggle. The overview of this dataset is, it has 27820 samples with 12 features. Download the dataset from the link provided.

Preprocessing

Data preprocessing is a critical step in the artificial intelligence process. Analysis of data without careful consideration can yield misleading results. For this reason, the representativeness and quality of the data should first be ensured before conducting the experiment. Preprocessing tasks include data cleaning such as identification and outlier removal, data integration, data transformation generating new features, and data reduction. The deliverable of the data preprocessing task is a new training set that ultimately leads to improved classification performance and reduced classification time. This is due to the reduced dimensionality of the data, which allows the learning algorithm to work faster and more efficiently. In some cases, it can increase the accuracy of future classifications. For others, the result is a more compact and interpretable representation of the target concept.

Feature Extraction

Tweets expressing suicidal thoughts do not have a pre-defined semi-fixed lexical-syntactic pattern. Therefore, it requires the use of hand engineering and analysis of a set of features, rather than embedding sentences or words in a supervised setting using deep learning models such as convolutional neural networks (CNNs). The proposed method makes use of the following set of features for classification. An RNN implementation of a tree-structured approach to extracting phrase semantics has been used for text classification with promising results in previous studies. However, the developed model suffers from the problem of gradient vanishing and exploding, so it takes a long time to develop a text tree structure for handling long tweets. Two forms of RNN-based design approaches have been created: LSTM and GRU. Both contain gating mechanisms that handle the limitations of RNNs. It incorporates "forget" gates that allow the network to encapsulate long-term relations without encountering the vanishing gradient problem.

Algorithm:
Begin
    Initialize the features population
    Calculate the feature fitness and weights

Select global and local leader
While (termination criteria is not satisfied) do
Updated the position of spider based female (Sf)
Updated the position of spider based male cooperative operators (Mf)
Perform matting processing
Termination criteria
Ite=iter + 1
End
Optimization solution
Stop

Feature selection

In the process of machine learning model development, only a few variables in the data set are useful for model construction, and the remaining features are redundant or irrelevant. Inputting a dataset with all these redundant and irrelevant features can negatively impact and degrade the overall performance and accuracy of the model. Therefore, it is very important to identify and select the most relevant features from the data, and to remove irrelevant or less important features by using machine learning feature selection.

```
Algorithm
start
feature_name = X.columns.tolist()
# calculate the correlation with y for each feature
        for i in X.columns.tolist():
         cor = np.corrcoef(X[i], y)[0, 1]
          cor_list.append(cor)
end
  # replace NaN with 0
          cor_list = [0 if np.isnan(i) else i for i in cor_list]
# feature name
          cor_feature = X.iloc[:,np.argsort(np.abs(cor_list))[-num_feats:]].columns.tolist()
 # feature selection? 0 for not select, 1 for select
        cor_support = [True if i in cor_feature else False for i in feature_name]
        return cor_support, cor_feature
        cor_support, cor_feature = cor_selector(X, y,num_feats)
        print(str(len(cor_feature)), 'selected features')
stop
```

Classification using RNN

The classification method is used to classify the best feature to predict the suicide case.RNNs generate feature classifications by convolving different subregions of a dataset with pre-trained kernels. In addition, nonlinear activation functions such as sigmoid, tanh, and linear correction functions can be applied. This method of reducing the amount of calculation is a data set, in which a feature region is selected, and the largest value in it is selected as a representative. It is used with a traditional fully-connected RNN and is usually used in the output stage.

Algorithm

Step 1: Initialization of the dataset
Step 2: To collect the dataset
Step3: Then, dataset are pre-processed
Step 4: The dataset under cleaning and data reduction process by using feature extraction.
Step 5: The trained features are selected by feature selection process.
Step 6: The proposed algorithm Recurrent Neural Network (RNN) is evaluated suicide prediction performance.
Step 5: finally produce the optimized result.

The generalized neural network different layers are represented as follows:

$$E[y|\boldsymbol{x}] = \frac{\int_{-\infty}^{\infty} y \cdot f(\boldsymbol{x}, y) \cdot \mathrm{d}y}{\int_{-\infty}^{\infty} f(\boldsymbol{x}, y) \cdot \mathrm{d}y}$$

RESULT AND DISCUSSION

The proposed system Recurrent Neural Network (RNN) for data processing is implemented for detecting suicide. suicide detection includes dataset, preprocessing, feature extraction, feature selection and classification. This describes a method for detecting the suicide case the algorithm is described for the detection of suicide.

Accuracy Performance

shows the analysis of accuracy level performance on the proposed algorithm RNN comparing with other algorithm.

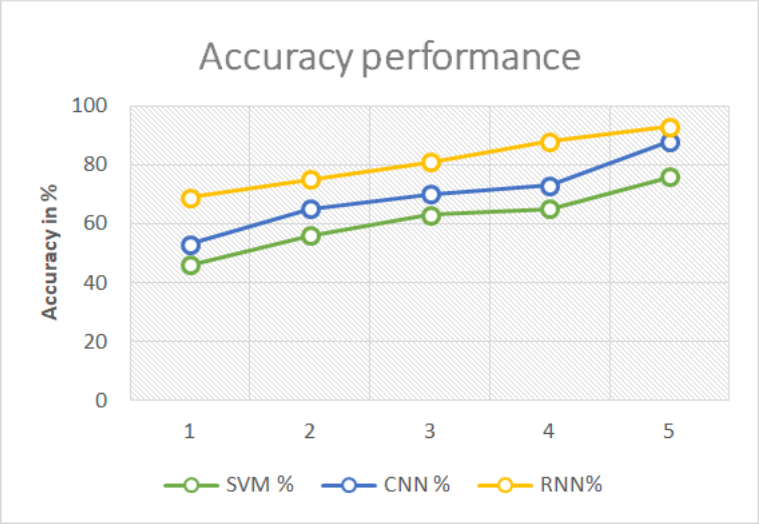

Fig 2: Analysis of accuracy performance

Figure 2 shows the analysis of accuracy performance in the proposed algorithm RNN. The existing system Support Vector Machine (SVM) provides 76 %, and CNN 88 %, and then, the proposed algorithm RNN provides 93% accuracy performance.

Prediction performance

Shows the analysis of prediction level performance on the proposed algorithm RNN comparing with other algorithm.

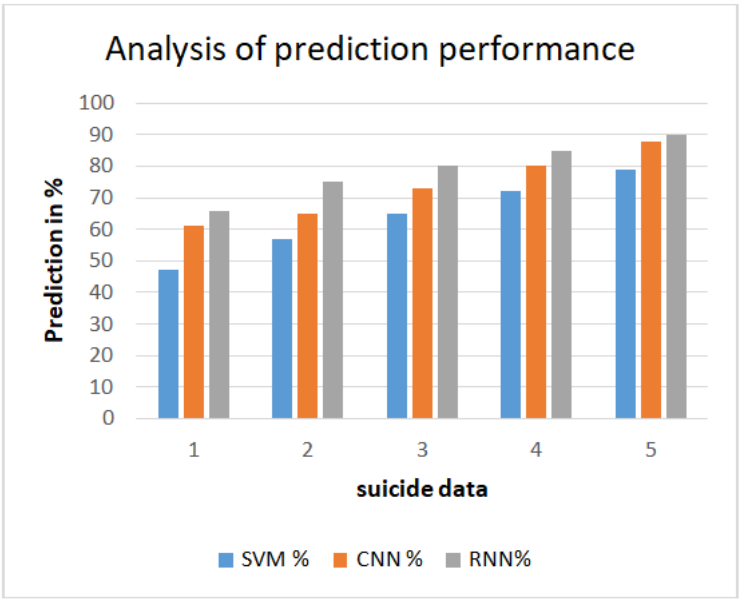

Fig 3: Analysis of prediction performance

Fig 3 shows the analysis of prediction level performance in the proposed algorithm RNN. The existing system Support Vector Machine (SVM) provides 79 %, and CNN provides 88%, and then the proposed algorithm RNN provides the prediction performance of 90%.

Time complexity performance

shows the analysis of time complexity level performance of the proposed algorithm RNN comparing with other algorithm.

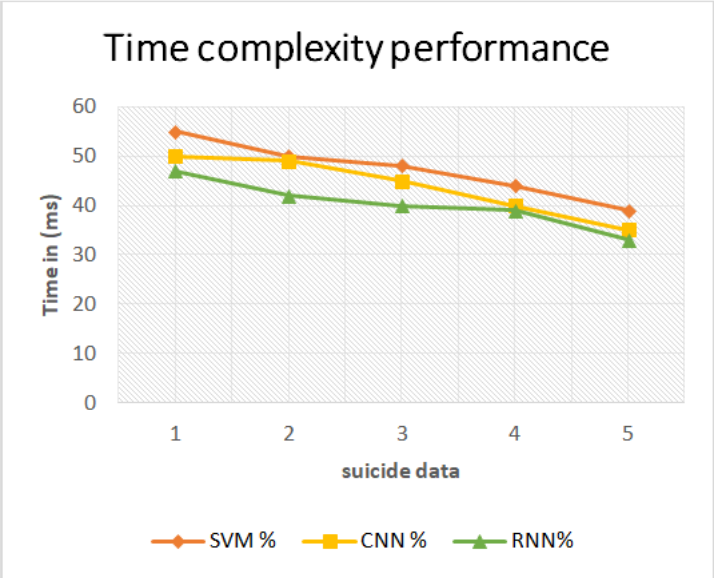

Fig 4: Analysis of Time complexity

Fig 4 shows the analysis of time complexity performance of the proposed algorithm RNN. The existing system Support Vector Machine (SVM) produces 39 ms, and CNN produces 35 ms, and then, the proposed algorithm RNN produces lowest time performance of 33 ms.

CONCLUSION

Suicide is one of the deadliest forms of human suicide, killing approximately 1 million people every year. Identification by depressive activity in humans is important given the current state of medicine. Used to find areas of the body that have developed malignant tumors. Data processing techniques, such as noise reduction, feature extraction, identification of damaged regions, and comparison with medical history data from suicide cases, were used to identify affected brain parts in humans. This ultimately results in an increase in quality. Next, the dataset is classified using recurrent neural network techniques.

It remains an open question whether predictive models and risk assessment tools can be applied to suicide prevention. The main studies in the three recent systematic reviews provided no evidence for clinical practice, but neither compared the model with current methods nor considered high rate values. Reviews themselves are limited and overstate their conclusions. Future work will not continue to develop new models in isolation, but rather investigate the added benefit of using these tools to support clinical decision making rather than replacing them. We need to move to real-world clinical evaluation.

REFERENCES

1. Lin GM, Nagamine M, Yang SN, Tai YM, Lin C, Sato H. Machine learning based suicide ideation prediction for military personnel. IEEE J Biomed Health Inform. Jul 2020;24(7):1907-16. doi: 10.1109/JBHI.2020.2988393, PMID 32324581.
2. Ren F, Kang X, Quan C. Examining accumulated emotional traits in suicide blogs with an emotion topic model. IEEE J Biomed Health Inform. Sep 2016;20(5):1384-96. doi: 10.1109/JBHI.2015.2459683, PMID 26208372.
3. Wang R, Yang BX, Ma Y, Wang P, Yu Q, Zong X et al. Medical-level suicide risk analysis: A novel standard and evaluation model. IEEE Internet Things J. 2021;8(23), Dec 1:16825-34. doi: 10.1109/JIOT.2021.3052363.
4. Xu Z, Zhang J, Zhang Q, Xuan Q, Yip PSF. A comorbidity knowledge-aware model for disease prognostic prediction. IEEE Trans Cybern. Sep 2022;52(9):9809-19. doi: 10.1109/TCYB.2021.3070227, PMID 33961578.
5. Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. IEEE Trans Bio Med Eng. Sep 2004;51(9):1530-40. doi: 10.1109/TBME.2004.827544, PMID 15376501.
6. Chadha A, Kaushik B. A survey on prediction of suicidal ideation using machine and ensemble learning. Comput J. 2021;64(11):1617-32. doi: 10.1093/comjnl/bxz120.
7. Baek J-W, Chung K. Context deep neural network model for predicting depression risk using multiple regression. IEEE Access. 2020;8:18171-81. doi: 10.1109/ACCESS.2020.2968393.
8. Seal A, Bajpai R, Agnihotri J, Yazidi A, Herrera-Viedma E, Krejcar O. DeprNet: a Deep Convolution Neural Network Framework for Detecting Depression Using EEG. IEEE Trans Instrum Meas. 2021;70:1-13:Art no. 2505413. doi: 10.1109/TIM.2021.3053999.
9. Qassem M, Constantinou L, Triantis IF, Hickey M, Palazidou E, Kyriacou PA. A method for rapid, reliable, and low-volume measurement of lithium in blood for use in bipolar disorder treatment management. IEEE Trans Bio Med Eng. Jan 2019;66(1):130-7. doi: 10.1109/TBME.2018.2836148, PMID 29993417.

10. Lejeune A, Le Glaz A, Perron PA, Sebti J, Baca-Garcia E, Walter M et al. Artificial intelligence and suicide prevention: a systematic review. Eur Psychiatry. 2022 Feb 15;65(1):1-22. doi: 10.1192/j.eurpsy.2022.8 [Epub ahead of print]. PMID 35166203, PMCID PMC8988272.
11. Halsband A, Heinrichs B. AI, suicide prevention and the limits of beneficence. Philos Technol. 2022;35(4):103. doi: 10.1007/s13347-022-00599-z.
12. Prince MC, Srinivas LNB. A review and design of depression and suicide detection model through social Media Analytics. Proceedings of the international conference on deep learning, computing and intelligence. Adv Intell Syst Comput. 2022;1396. doi: 10.1007/978-981-16-5652-1_40.
13. Cao L, Zhang H, Wang X, Feng L. Learning Users Inner Thoughts and Emotion Changes for Social Media based Suicide Risk Detection. IEEE Trans Affect Comput:1-. doi: 10.1109/TAFFC.2021.3116026.
14. Naseem U, Khushi M, Kim J, Dunn AG. Hybrid text representation for explainable suicide risk identification on social Media. IEEE Trans Comp Soc Syst. 2022:1-10. doi: 10.1109/TCSS.2022.3184984.
15. Ma Y, Cao Y. Dual Attention based Suicide Risk Detection on Social Media IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China. 2020. p. 637-40. doi: 10.1109/ICAICA50127.2020.9182380.