



Daffodil
International
University

Predicting Suicidal behavior

Due to

COVID-19: An ANN Approach

Submitted By

Name: MD. Rakibuzzaman

Id: 181-35-347

Sec: A

Submitted to

Dr. Imran Mahmud

Associate Professor & Head

Department of Software Engineering
Daffodil International University


A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in
Software Engineering

© All right Reserved by Daffodil International University

APPROVAL

This thesis/project/internship entitled on *“Predicting Suicidal Behavior Due to COVID-19: An ANN Approach”*, submitted by *Md. Rakibuzzaman* (ID: 181-35-347) to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS



Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University



External Examiner

Abu Shamim Aminur Razzaque
Director
Computer Ease Limited



Internal Examiner 1

SK. Fazlee Rabby
Lecturer
Department of Software Engineering
Daffodil International University



Internal Examiner 2

Md. Rajib Mia
Lecturer
Department of Software Engineering
Daffodil International University

DECLARATION

I hereby declare that I have taken this thesis under the supervision of **Dr. Imran Mahmud, Associate professor & head, Department of Software Engineering, Daffodil International University**. I also declare that neither this thesis nor any part of it has been submitted elsewhere for award of any degree or diploma.



Md. Rakibuzzaman

ID: 181-35-347

Batch: 25th

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

Certified by:



Dr. Imran Mahmud

Associate professor & head

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

Abstraction

Covid-19 has been a major worry over the world in recent years. At the time of this epidemic, psychological problems as well as suicidal issues are being noticed among the people. The main research of this paper will be mainly on the prediction purpose of Suicidal Behavior. Most of them are young and the age range is 18-35 years. This research is based on the data of “Achol Foundation” which contains 2077 columns and 30 features. Using the Deep Learning Concept- Artificial Neural Network model used to predict the suicidal behavior of a single person. For this research, the parameters have been tuned by cleaning the data from complete data through data analysis. Data cleaning has been done through Excel and Python libraries and modeling has been done at Matlab. Using the ANN model, the accuracy score is 79.42% which has been able to give a really good result. In the future, steps may be taken to arrange online consultancy based on this prediction of who is going to attempt suicide.

Table of Contents

Approval	i
DECLARATION	ii
Abstraction	iii
Table of Contents	iv
Chapter 1: Introduction	1
1 Sample of suicidal case in Bangladesh.....	1
1.1.1 Case 1.....	1
1.1.2 Case 2.....	1
1.2 Most Important Factor for Suicide Tendency.....	2
1.3 Research Questions.....	2
Chapter 2: Background Study	3
2.1 Literature Review.....	3
Chapter 3: Research Methodology	5
3.1 Data Collection.....	5
3.2 Measurements items.....	5
3.3 Data Cleaning.....	6
3.3.1 Data Preprocessing.....	6
3.3.2 Data Labeling.....	8
3.3.3 Data Categorization.....	9
3.4 ANN Model Train	10
3.5 Cross Validation.....	11
3.6 Important Factors.....	11
Chapter 4: Result & Discussion	13

4.1 Hyperparameters.....	13
4.2 Hyperparameters tuning.....	14
4.3 Classification Result.....	15
4.4 Validation Result.....	16
4.5 Confusion Matrix.....	16
4.6 Precision, Recall and F-Score.....	18
Chapter 5: Conclusions and Recommendations.....	19
Reference:	20

Chapter 1

Introduction

According to the Pan American Health Organization, suicide prevention must be focused after 18 months of the COVID-19 epidemic (PAHO) [1]. “Suicide Is Not Just a US Problem, It's a Global Issue”. Suicide is also a worldwide issue. Suicide is the second-leading cause of mortality among 15- to 29-year-olds, according to the World Health Organization, with 75 percent of suicides occurring in low- and middle-income nations [2]. This problem actually is not only a global problem there is a lot of examples in Bangladesh also. This study aims two special cases- I) Based on ecological research, the incidence of suicide behavior and risk variables and another is II) Suicide stresses connected to the COVID-19 epidemic, based on case studies [3].

Here are some of the cases involving COVID-19's suicidal behavior. I gathered this data from a variety of sources. They showed many suicidal incidents that occurred as a result of Covid-19 in the study [4].

1.1 Sample of Suicidal Case in Bangladesh

- 1.1.1 **Case 1:** On April 12, a 35-year-old mother of five from Cox's Bazar attempted suicide by hanging, but one of her sons saved her by enlisting the support of her neighbors. Because of the lockdown, her husband lost his job, and they were also unable to receive relief items from the local government. The mother couldn't endure seeing her starving children's faces and decided to commit herself in order to give more food for them (Campus Today 2020).
- 1.1.2 **Case 2:** On April 10, a female adolescent (aged 10) from Sirajgonj's Belkuchi municipality committed suicide (by hanging herself) after her father reprimanded her for begging for food. Because of the shutdown, the girl's father had to close his modest loom factory, leaving the family penniless. The entire family had gone hungry for a few days and had been refused any financial assistance from the local government (Kaler Kantho 2020a).

1.2 Most Important Factor for Suicide Tendency

- To findout the factor we are using Artificial Neural Network(ANN) model.
- Neural Networks come into picture when we have a high amount of output classes and high amount of data to support the performance of the model. In this case our data number is large and for that we will use ANN from the concept of deep learning model.
- And finally, we can make perfect predictions of suicidal behavior based on important variables.

1.3 Research Questions

Sample of Questionire and Dataset:

- Do you have moderate sleep? (7-8 hours)
- How much time do you spend on social media?
- Have you ever suffered from depression? (Being depressed most of the time, loss of interest and interest in activities that you enjoy, sleep may be abnormally low or high, slow in work and thinking, negative thinking or feeling responsible, lack of decision or attention in everything) And planning and attempting suicide if very severe)

Chapter 2

Background Study

Pandemic panic is now worldwide. Therefore, various research papers related to this epidemic have been being prepared for the last two years. We have come to the conclusion by reading different types of research papers. Literature reviews of various such papers have been discussed here. The study's goal was to look at the rate of clinical depression in the Greek adult population aged 18–69, as well as the changes in anxiety, discomfort, and suicidal thoughts [5]. To do this they have basically used the statistical method. They have given a result of the effect of doing from a certain age during lockdown. From there they extracted a sample of suicidal people.

"Suicide prevention during the COVID-19 epidemic" is the topic of this study. The study found a number of suicidal explanations for the epidemic and chose to take preventative measures [6].

They explored an upsurge in suicide after an initial reduction during the COVID-19 epidemic in Japan in their research [7].

Anxiety, sadness, and suicidal behavior are all linked to sleep disruptions. Suicidal thoughts, suicide attempts, and death are all linked to sleep disturbances [8].

2.1 Literature Review

Konstantinos N Fountoulakis a,* , Maria K. Apostolidou b are the author of the research paper which is “Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece” (2020). The main variables of this papers are Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece [10].

They have used Machine Learning algorithms such as- Linear Regression. Also they implemented statistical analysis using Descriptive statistics- Chi square tests, Anova testing. The dataset is downloaded from different online platform and anonymously.

“Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review” paper is published on 09 November, 2020 by Gema Castillo-Sánchez, Gonçalo Marques [11]. The dataset of this paper collected by General social networks (GSNs), Twitter (10/16, 62.5%), microblogs (3/16, 18.75%),Others (2/16, 12.5%), , Facebook, Instagram, Tumblr, and Reddit (1/16, 6.25%). They used

different machine learning algorithms like Support Vector Machine, Linear Regression, Naïve Bayes, K-Nearest Neighbour, Random Forest.

Mohammed A Mamun 1,2 published an article on 3 June, 2021. The name of this paper is “Suicide and Suicidal Behaviors in the Context of COVID-19 Pandemic in Bangladesh: A Systematic Review” [12]. COVID-19 and psychological impact, pandemic and suicide, prevalence and risk factors, suicide and self-harm, suicide deaths in Bangladesh, hanging in Bangladesh are the main variables of this research paper. The dataset was collected from databases like PubMed, Scopus, PsycINFO, Web of Science, CINAHL.

There is another article which was published on March 11, 2021 which is a most recent paper. The topic of this research paper is “Prediction of acute suicidal ideation in young adults using multi-dimensional scales: A graph neural network approach” [13] published by Kyu Sung Choi, Byung Hoon Kim.

The main variables of this paper is COVID-19, deep learning, depression, suicide, graph neural network. The data acquired from validation, questionnaires and other clinical data, anonymized prior to combining the data from the four institutions. The analysis section of this paper is completed by using machine learning algorithms such as- Logistic regression with LASSO, Support Vector Machine, GIN_ SMOTE, Ensemble of GINs (Mind WatchNet).

This paper is being done on a very recent topic so there are very few such papers online. This paper has been written after reading from various online research papers and online famous news site.

Chapter 3

Research Methodology

This segment will discuss the entire research pipeline. From data collection to data modeling, the complete cycle works will be described and the modeling architecture will be shown. This research takes more time as the data cleaning part is complicated. The data set is then prepared for deployment by performing some important tasks such as analysis from the dataset. The pipeline figure for the full data science research topic given below:

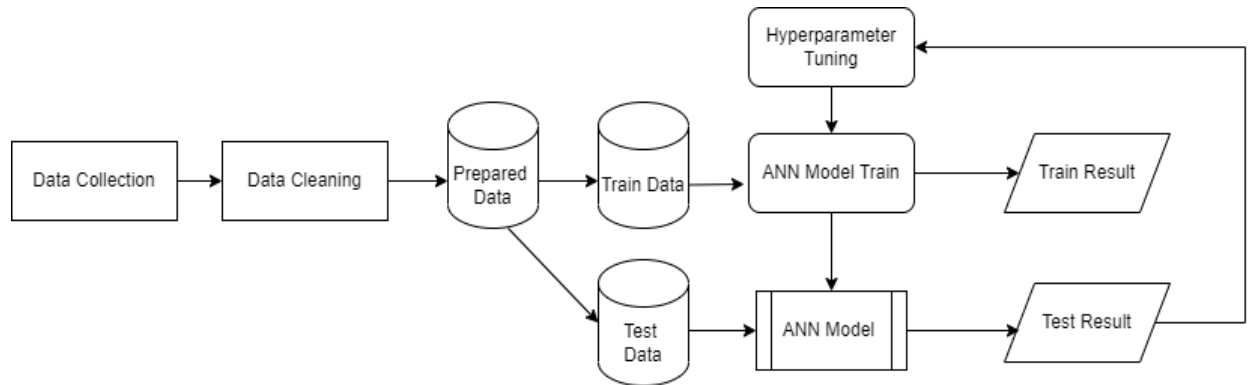


Fig.2 Pipeline of Model development

3.1 Data Collection:

- This dataset is collected by “Achol Foundation for mental health”.
- Sample size of the research will be Youth: 18-35 years.
- Link of the dataset:
https://github.com/roman1117459/thesis_all_final/blob/master/converted.xlsx

3.2 Measurements items

- This data set has 2077 rows and 30 columns.

Fig. 3 Checking null values

There we check the importance of various features and if there is excess value, I remove those columns. Also features that were above null value 40-50% are removed. I have implemented this code for find out the most null values from those features which show me the below figure :

```
null_cols = (dataset.isnull().sum() / len(dataset)).sort_values(ascending=False)[:20]
```

```
null_cols
```

```
Out[5]: Yourthoughtson_the_mentalhealth_ofyoungpeople      0.999037
        Areyouaware_of_theAreaFoundation                  0.998556
        Howpositivedoyou_think_yourlifeis                 0.998556
        ever_consulted_mentalhealthprofessional           0.998556
        opinion_for_mhp                                    0.998556
        have_familymember_mhp                             0.998556
        References                                         0.573905
        favor_of_openinganeducational_institution         0.024555
        Whatdidyoudotocopewithstress_incorona             0.024555
        have_everthought_of_suicide_incorona              0.024555
        mostresponsible_factor_epidemic                   0.024555
        kindofproblems_facing_inCorona                    0.024555
        Howisyour_coronary_stress_more_than_before        0.024555
        doyouknow_psychitrist_solveyour_mhp              0.002407
        Ever_thought_of_suicide                           0.001444
        consulted_a_psychiatrist                          0.001444
        every_harmedurself_depression                     0.001444
        with_whom_share_depression                       0.001444
        do_wt_when_suffer_depression                      0.001444
        biggest_cause_mhp                                 0.001444
        dtype: float64
```

Fig. 4 Checking Null values Percentage

Finally I have removed the columns which contains maximum null values because those columns have no strong impact for this analysis. I filled the data using median because all the features were in object type. I checked the data to see if all the data was filled properly. In addition, the features that had excessive variation, meaning that it was not possible to categorize, have been removed. Because those features were not machine readable, meaning they were not significant. In this way basic data analysis has been done very easily through Python.

3.3.2 Data Labeling:

"Achol Foundation" collected data in CSV format through Data Online. Because the data is in CSV format, it is easy to sort the data, name the features of the data through Microsoft Excel. At this stage the names of the datasets were all in the form of questionire. So the first challenge was to shorten the questions of the features, remove space and the numerical keyword with different features of the dataset. Because the machine will not be able to read the dataset that we will use in the program. So the questions have been shortened. Below is a table of samples short names from the feature question –

Shortened Features Column	Questionaire
age	1. Your age
occupation	3. Occupation?
married	4. Marital status?
sleep	5. Do you have moderate sleep? (7-8 hours)
socialmediatime	. How much time do you spend on social media?
depression	. Have you ever suffered from depression?
suffforsharesomething	. Suffering from emotional instability get someone to share?

Table 1 Shortening questions

Microsoft Excel also completes other small tasks such as removing dot (.) From features and column values, categorizing category features.

3.3.3 Data Categorization

I have used Matlab for data modeling. The coding part of which is completely in the suicide.m file and there we convert the categorical data to numeric data through One hot encoding with the following code.

```
data = convertvars(data,categorical_input_names,'categorical');  
  
for i = 1:numel(categorical_input_names)  
  
name = categorical_input_names(i);  
  
oh = onehotencode(data(:,name));  
  
data = addvars(data,oh,'After',name);  
  
data(:,name) = [];  
  
end
```

I basically use that code to categorize the data. So that every categorical data is now in numeric format and also it is machine readable.

3.4 ANN Model Train:

I used Artificial Neural Network model for suicide attempt prediction. It is better to use deep learning concept for prediction than other models because ANN model can give good prediction if dataset is big. I used ANN because it is an industrial data and such data will increase in the future. The ANN model architecture is shown below:

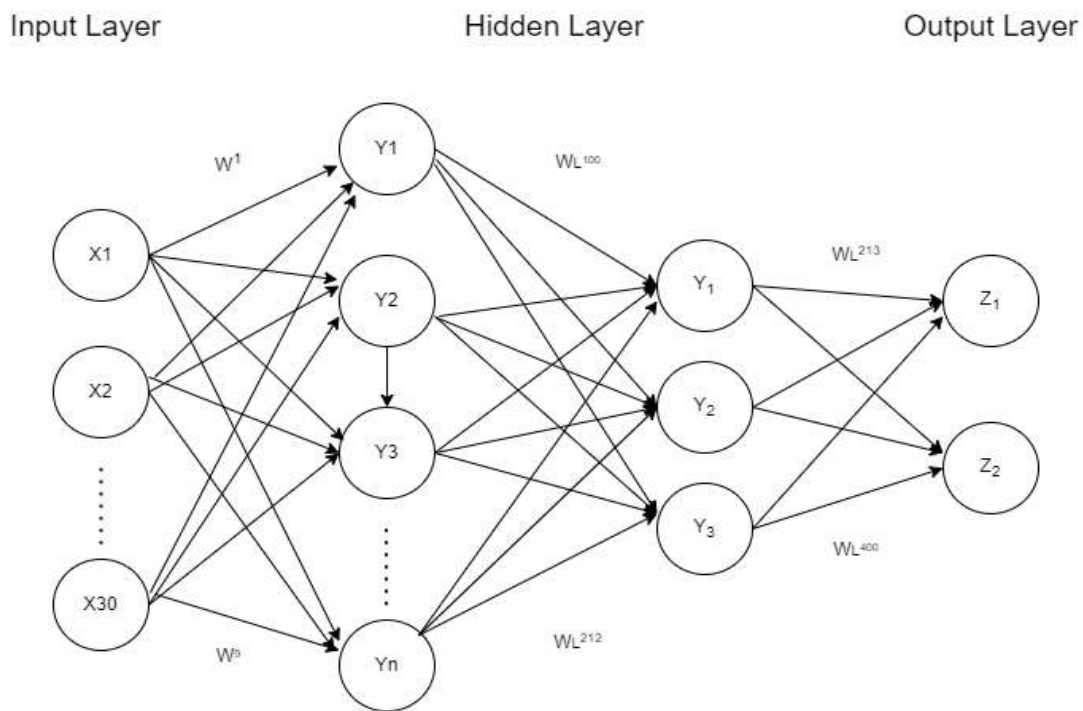


Fig. 1 ANN Architecture

The model transmits information from one layer to another through its weight and makes output predictions. All these hidden layer parameters can be controlled through tuning. This results in better results and better modeling. I will discuss model tuning in the next chapter.

3.4 Cross Validation:

The model must be cross-validated to avoid any kind of overfitting and underfitting problems. I used 70% of the data for training and the remaining 30% for data testing as the data in this data set is quite good. Since we did the coding part through Matlab, we did the manual crossvalidation. The code is given below-

```
data_size = size(data,1)
training_data_size = floor(0.7*data_size)
validation_data_size = floor(0.15*data_size)
test_data_size = data_size - training_data_size - validation_data_size
idx = randperm(data_size);
training_idx = idx(1:training_data_size);
validation_idx = idx(training_data_size+1:training_data_size+validation_data_size);
test_idx = idx(training_data_size+validation_data_size+1:end);
training_data = data(training_idx,:);
validation_data = data(validation_idx,:);
test_data = data(test_idx,:);
```

The above code uses “randperm” library as if the data is used randomly during cross validation. It's a lot like K Fold cross validation. But here I have arranged the code completely manually so that the work can be done very perfectly in Matlab.

3.4 Important Factors:

From the features obtained from the dataset here, we can identify the important factors for suicidal behavior through “corr” function. Here are some of the 19 most important features that are responsible for suicidal behavior.

Column	Correlation Value
haveeverthoughtofsuicideincorona_No	0.478106
depression_Yes	0.466617
everyharmedurselfdepression_Yes	0.426027
suffforsharesomething_Yes	0.256985
mhpinterferedailytask_Lots	0.285154
coronarystressmorethanbeforeornot_Much more than before	0.243850
mhpinterferedailytask_No	0.215477
withwhomsharedepression_noone	0.199225
consultedapsychiatrist_Yes	0.144492
coronarystressmorethanbeforeornot_The same	0.140418
socialmediatime_Poor	0.105145

Table 2 Correlation value

Chapter 4

Result & Discussion

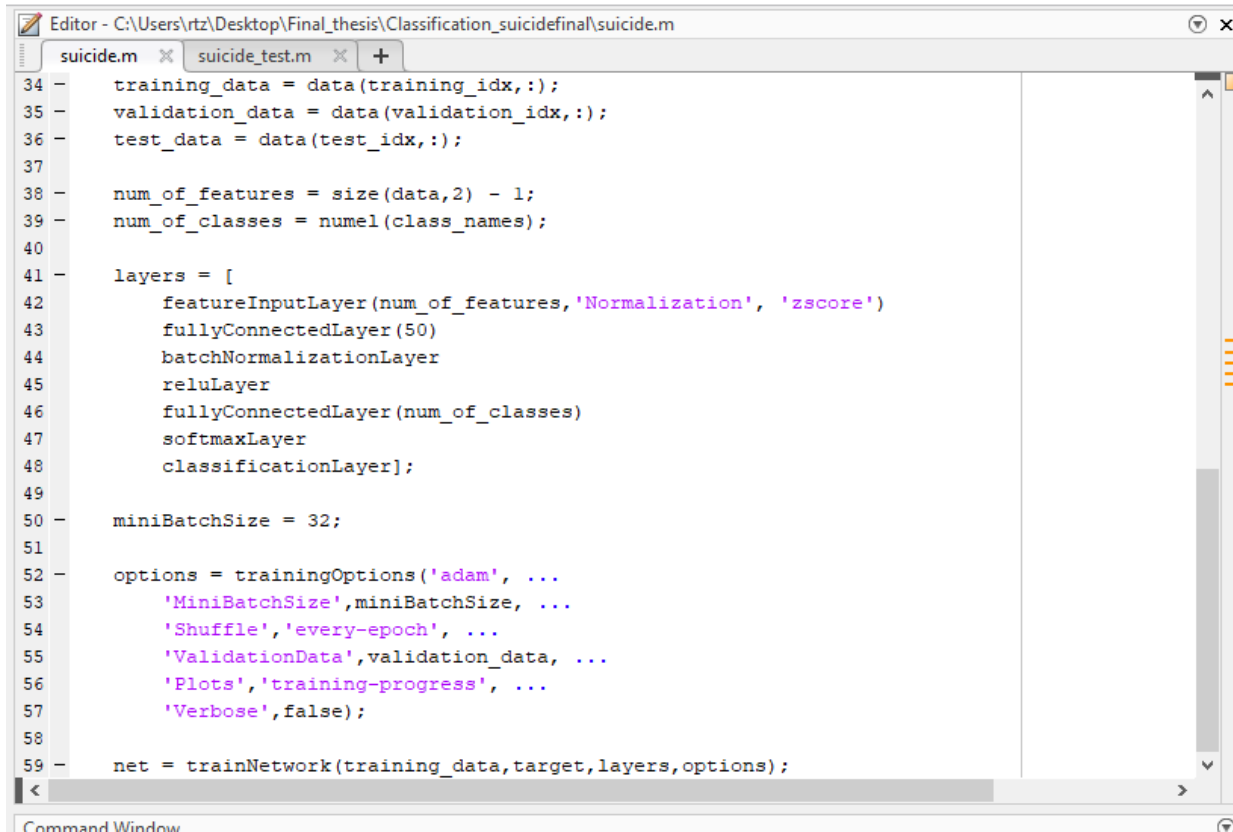
In this section we will mainly discuss the results of our research and performance measurement. Here we will basically show the performance of the model and discuss the parameters that we have used for the prediction of suicidal behavior.

4.1 Hyperparameters:

Some parameters for neural network must be given before model training. These parameters determine the timing and performance of model training. Which can be reduced or increased by tuning the accuracy. I have declared different layers function in batch variables and different loss function, batch size, epoch etc. in options variable. I used the following parameters and modified the code in different ways to get the highest score.

- Normalization: “rescale-symmetric”.
- Loss function : adam
- Mini batch size : 32
- Fully Connected Layer : 50
- Shuffle : “every-epoch”
- Verbose : False

Attached is a screenshot of the Matlab coding file of my hyper parameter tuning below.



```
Editor - C:\Users\rtz\Desktop\Final_thesis\Classification_suicidefinal\suicide.m
suicide.m  x  suicide_test.m  x  +
34 - training_data = data(training_idx,:);
35 - validation_data = data(validation_idx,:);
36 - test_data = data(test_idx,:);
37
38 - num_of_features = size(data,2) - 1;
39 - num_of_classes = numel(class_names);
40
41 - layers = [
42     featureInputLayer(num_of_features,'Normalization','zscore')
43     fullyConnectedLayer(50)
44     batchNormalizationLayer
45     reluLayer
46     fullyConnectedLayer(num_of_classes)
47     softmaxLayer
48     classificationLayer];
49
50 - miniBatchSize = 32;
51
52 - options = trainingOptions('adam', ...
53     'MiniBatchSize',miniBatchSize, ...
54     'Shuffle','every-epoch', ...
55     'ValidationData',validation_data, ...
56     'Plots','training-progress', ...
57     'Verbose',false);
58
59 - net = trainNetwork(training_data,target,layers,options);
Command Window
```

Fig. 5 Hyper parameters

4.2 Hyperparameters tuning :

At first my accuracy was 76.85%. Next I tuned the number of layers, loss function, normalization etc. in different ways. And lastly, when I normalize the value of "rescale-symmetric" and fully connected layers to 50, my accuracy increases to 79.42%. So tuning the parameters is very important to build a good model.

4.3 Classification Result :

I have used the ANN model to predict suicidal behavior from datasets. Accuracy comes to 79.42% after our model training after coding party matlab. Which is really capable of giving very good predictions. Here attaching a screenshot of performance analysis during training the model:

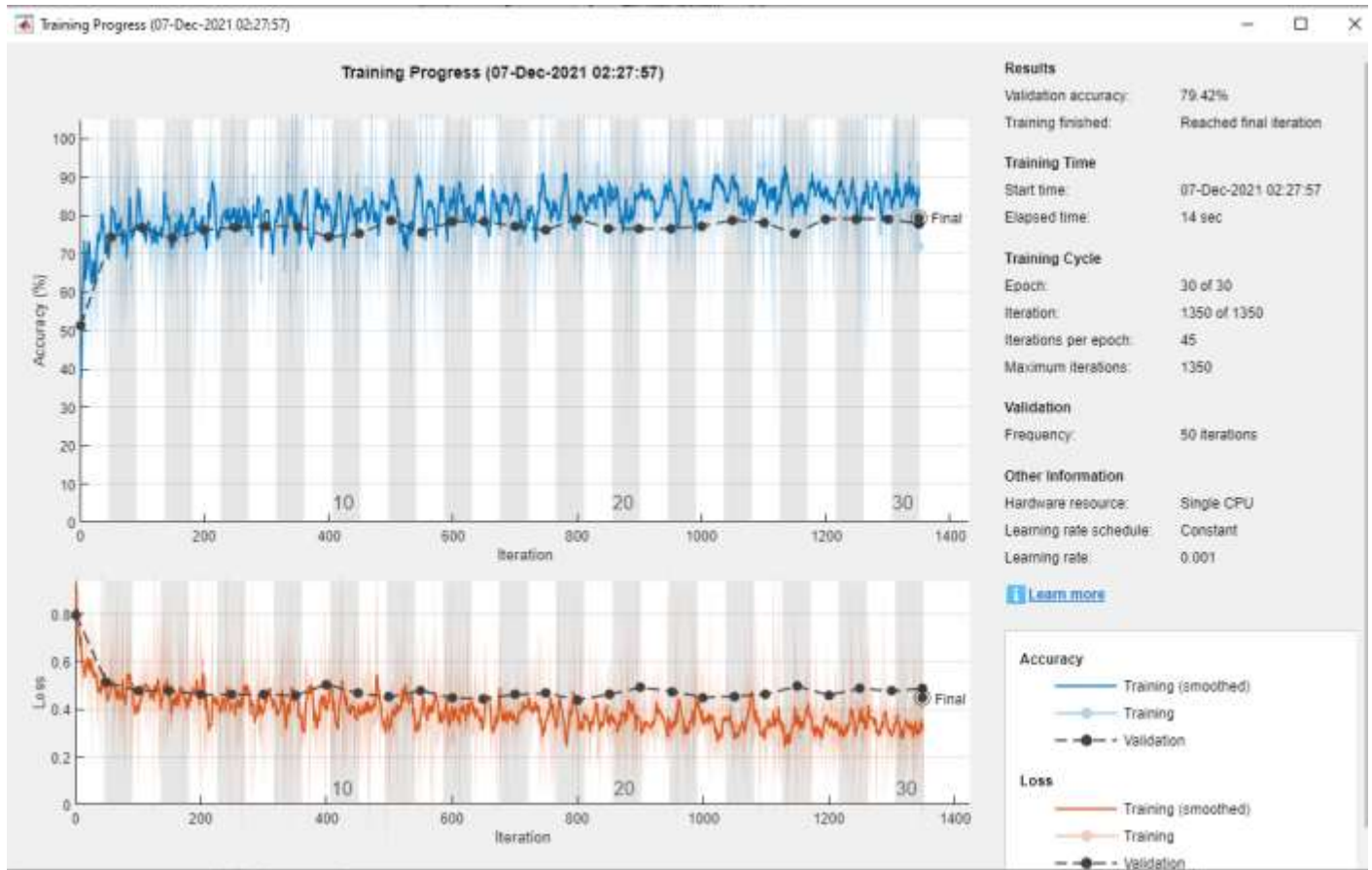


Fig. 6 Performance Analysis

In the above screenshot we can clearly see that for model training I have very nicely determined the accuracy and loss of our dataset through 30 epochs and 50 iterations. Here it is seen that with iteration the loss decreases and the accuracy increases. Which means my model worked well and the accuracy is 79.42

which is a model capable of giving really good predictions. Now my task is to check whether the model is overfitting or underfitting by measuring its performance.

4.4 Validation Result :

In addition, after running the model on the validation set, the result is 0.7125. Accuracy is reduced here because our model is not aware of the complete dataset now. Then 71% accuracy really gave a very satisfactory performance. The file we stored in Matlab called suicide_test.m. Here we have plotted our model performance and Confusion matrix. Attached is a screenshot of the coding file below-

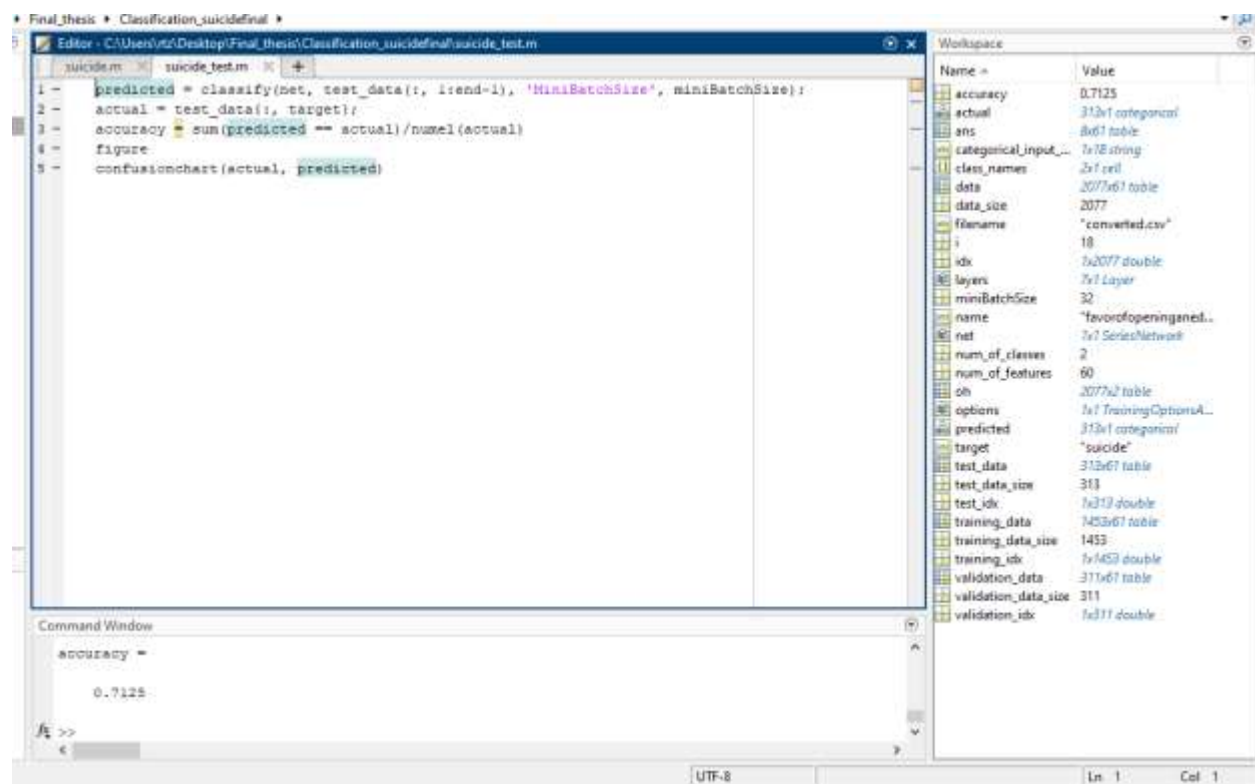


Fig. 7 Validation Code

4.5 Confusion Matrix :

It is easy to judge the good and the bad of any model through the Confusion matrix. Because from this the difference between Actual Class and Predicted class can be measured through the values of True Positive,

True Negative, False Positive, False Negative of the model. Here is the confusion matrix for our prediction model:

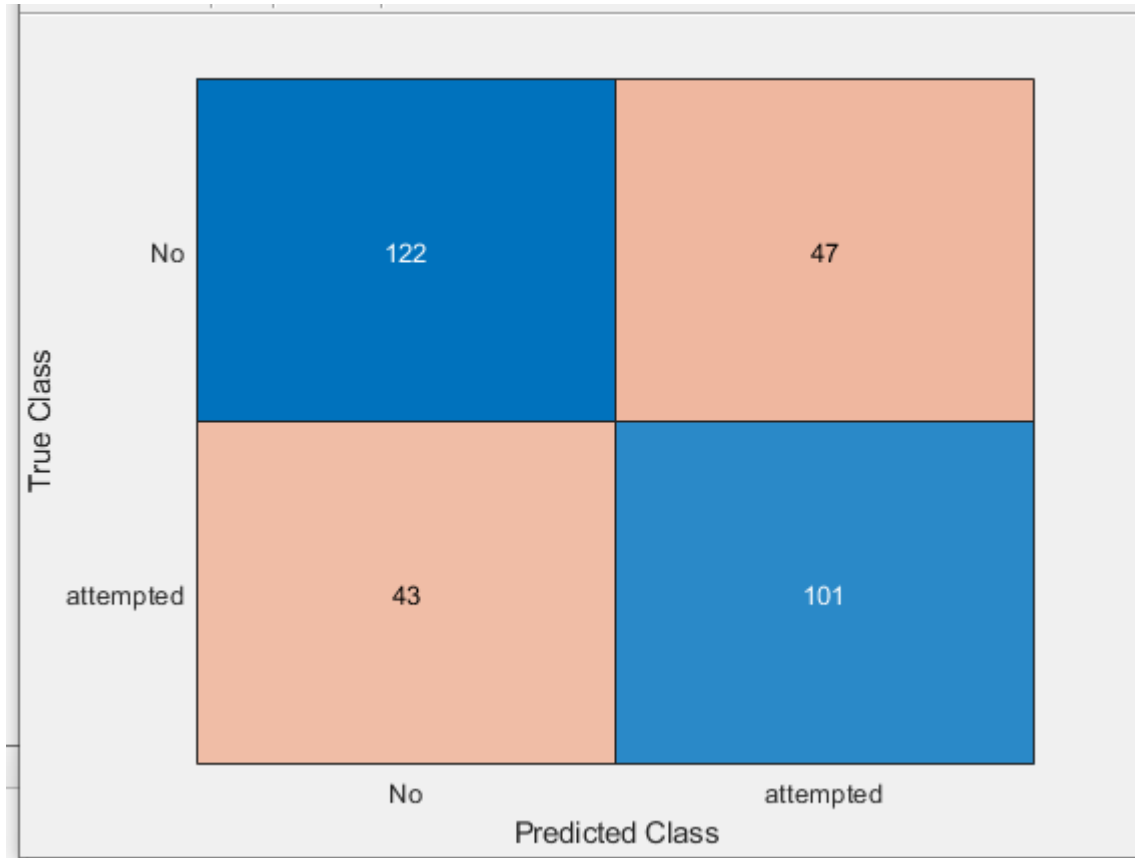


Fig. 8 Confusion Matrix

4.6 Precision, Recall and F-Score :

I generate all the values for my model in Matlab using the Confusion matrix and then extract the Precision, recall and F-score from those values using the manual formula. These values are given below- Model performance is measured by the result of the Confusion matrix based on the validation set.

- Recall Score : 0.70138
- Precision Score : 0.68243
- F-Score : 0.69178

If we take a closer look at the above values, we will see that there is no possibility of data implantation and over fit or underfit. Which is really a good model performance.

Chapter 5

Conclusions and Recommendations

This paper is based primarily on a Predictive Purpose of Suicidal Behavior for Pandemics. The number of datasets was adequate and the number of datasets will increase further in the future as it is industry based data. And for that I have analyzed and modeled a little differently without approaching other models. As a result, it is able to make about almost 80 percent predictions from the data without any difficulty. In most of the research, machine learning was originally compared between different algorithms but very few papers have been worked on using the deep learning concept. If data can be increased in the future and more work can be done with a new concept. As a result, accuracy will increase and it will be able to solve this suicidal behavioral problem in our country and in the world.

Reference:

1. Hossain MM, Sultana A, Purohit N. Mental health outcomes of quarantine and isolation for infection prevention: a systematic umbrella review of the global evidence. *Epidemiol Health*. 2020;42: e2020038. doi:10.4178/epih.e2020038
2. Mamun MA. The first COVID-19 triadic (homicide!)-suicide pact: do economic distress, disability, sickness, and treatment negligence matter? *Perspect Psychiatr Care*. 2020. doi:10.1111/ppc.12686
3. Tasnim R, Islam MS, Sujon MSH, Sikder MT, Potenza MN. Suicidal ideation among Bangladeshi university students early during the COVID-19 pandemic: prevalence estimates and correlates. *Child Youth Serv Rev*. 2020;119:105703. doi:10.1016/j. childyouth.2020.105703
4. Shigemura J, Ursano RJ, Morganstein JC, et al. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: mental health consequences and target populations. *Psychiatry Clin Neurosci* 2020 Apr;74(4):281e2. <https://doi.org/10.1111/pcn.12988>. Epub 2020 Feb 23
5. Lai J, Ma S, Wang Y, et al. Factors associated with mental health outcomes among health care workers exposed to Coronavirus Disease 2019. *JAMA Netw Open* 2020 Mar 2;3(3):e203976. <https://doi.org/10.1001/jamanetworkopen.2020.3976>.
6. Xiao H, Zhang Y, Kong D, et al. Social capital and sleep quality in individuals who self-Isolated for 14 days during the Coronavirus Disease 2019 (COVID19) outbreak in January 2020 in China. *Med Sci Monit* 2020;26:e923921. <https://doi.org/10.12659/MSM.923921>.
7. Perez-Fuentes et al., 2020 M.D.C. Perez-Fuentes, M.D.M. Molero Jurado, A. Martos Martinez, J.J. Gazquez Linares Threat of COVID-19 and emotional state during quarantine: positive and negative affect as mediators in a cross-sectional study of the Spanish population *PLoS ONE*, 15 (2020), Article e0235305
8. Tomljenovic et al., 2020 H. Tomljenovic, A. Bubic, N. Erceg It just doesn't feel right - the relevance of emotions and intuition for parental vaccine conspiracy beliefs and vaccination uptake *Psychol. Health*, 35 (2020), pp. 538-554

9. F. Ornell, J.B. Schuch, A.O. Sordi, et al. "Pandemic fear" and COVID-19: mental health burden and strategies *Braz J Psychiatry* (2020 Apr 3), 10.1590/1516-4446-2020-0008 pii: S1516-44462020005008201 [Epub ahead of print]
10. J. Shigemura, R.J. Ursano, J.C. Morganstein, et al. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: mental health consequences and target populations *Psychiatry Clin Neurosci*, 74 (4) (2020 Apr), pp. 281-282, 10.1111/pcn.12988 Epub 2020 Feb 23
11. J. Lai, S. Ma, Y. Wang, et al. Factors associated with mental health outcomes among health care workers exposed to Coronavirus Disease 2019 *JAMA Netw Open*, 3 (3) (2020 Mar 2), Article e203976, 10.1001/jamanetworkopen.2020.3976
12. S. Li, Y. Wang, J. Xue, et al. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users *Int J Environ Res Publ Health*, 17 (6) (2020), p. E2032, 10.3390/ijerph17062032