

**CORONA VIRUS CLASSIFICATION USING MACHINE LEARNING  
TECHNIQUE IN THE CONTEXT OF BANGLADESH.**

**By**

**DIPTA SAHA**

**ID: 193-25-833**

This report presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering.

Supervised By

**Md. Zahid Hasan**

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised By

**Dr. Sheak Rashed Haider Noori**

Associate Professor and Associate Head

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

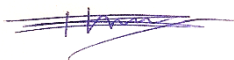
**DHAKA, BANGLADESH**

**JUNE, 2021**

## **APPROVAL**

This project titled “**Corona Virus classification using machine learning technique in the context of Bangladesh**”, submitted by Dipta Saha, ID: 193-25-833; to the Department of Computer Science and Engineering has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 03 JUNE, 2021.

## **BOARD OF EXAMINERS**



---

**Dr. Touhid Bhuiyan**

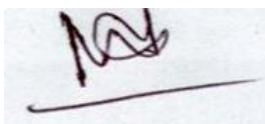
**Professor and Head**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

**Chairman**



---

**Dr. Md. Ismail Jabiullah**

**Professor**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

**Internal Examiner**



---

**Dr. Sheak Rashed Haider Noori**

**Associate Professor and Associate Head**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

**Internal Examiner**



---

**Dr. Shamim H Ripon**

**Professor**

Department of Computer Science and Engineering  
East West University

**External Examiner**

## DECLARATION

I hereby declare that, this project has been done by myself under the supervision of Md. Zahid Hasan, Assistant Professor, Department of CSE, Daffodil International University. I also declare that neither this project nor any part of this project consciously or unconsciously has been submitted elsewhere for award or any other degree.

### Supervised by:



---

**Md. Zahid Hasan**  
Assistant Professor  
Department of CSE  
Daffodil International University

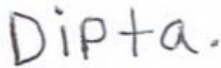
### Co-Supervised by:



---

**Dr. Sheak Rashed Haider Noori**  
Associate Professor and Associate Head  
Department of CSE  
Daffodil International University

### Submitted by:



---

**Dipta Saha**  
ID: 193-25-833  
Department of CSE  
Daffodil International University

## **ACKNOWLEDGEMENT**

First I express my heartiest thanks and gratitude to almighty God for His blessings which made possible for me to complete the final year project successfully.

I really grateful, thankful and also wish my indebtedness to my honorable teacher Md. Zahid Hasan, Assistant Professor, Department of CSE, Daffodil International University, Dhaka. To carry out this project the thing which helped me most is the deep knowledge, keen interest and helpful mentality of my supervisor in the field of “Machine Learning”. His endless patience, scholarly guidance, continual encouragement, energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would also like to express my heartiest gratitude to the almighty God and Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank my all course mate in Daffodil International University, who took part and help me in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents, family, friends and beloved persons.

## **ABSTRACT**

The covid-19 pandemic has spread to 221 countries and regions of the world, significantly affecting the global economy. Since the virus has been growing exponentially, even the developed countries have been unable to contain its spread. As a result, people are dying in the affected areas at an alarming rate. We can barely think how deadly and highly contagious disease will turn if it spreads from a moderate to a strong category in Bangladesh. This paper represent classification of corona virus based on symptoms and actions. The virus latches its spiky surface proteins to receptors on healthy cells mostly in lungs. The body does not recognize the pathogens so that immune response being triggered by the virus which can cause inflammation. Around 80% of people with corona virus recover without any special treatment within 2weeks mostly. For them, this is nothing but flu like symptoms. But those who are elder and having many more disease might feel difficulty in breathing, low level of oxygen in blood, lung injuries and pneumonia while effected by corona virus. Sometimes the virus can create life threatening situation as it clear the body's immune reaction and lowering the oxygen level in blood.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Approval	i-ii
Declaration	iii
Acknowledgement	iv
Abstract	v
<b>CHAPTER 1: INTRODUCTION</b>	
1.1 Introduction	1-2
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Question	3
1.5 Expected Output	3
1.6 Report layout	3-5
<b>CHAPTER 2: BACKGROUND</b>	
2.1 Introduction	6
2.2 Related Work	6-8
2.3 Research Summary	8
2.4 Scope of the Problem	8
2.5 Challenges	8-9

## **CHAPTER 3: RESEARCH METHODOLOGY**

3.1 Introduction	10
3.2 Research Subject and Instrumentation	11
3.3 Data Collection Procedure	11
3.4 Statistical Analysis	11-12
3.5 Implementation Requirements	12

## **CHAPTER 4: EXPERIMENTAL RESULT AND DISCUSSION**

4.1 Introduction	13
4.2 Experimental Result	13
4.3 Descriptive Analysis	13-18
4.4 Summary	18

## **CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY**

5.1 Summary of the study	19
5.2 Conclusion	19
5.3 Recommendation	19
5.4 Implication for further study	19



## **CHAPTER 6: CONCLUSION AND FUTURE SCOPE**

6.1 Conclusion	20
6.4 Future Scope	20
<b>REFERENCES</b>	21-22
<b>APPENDICES</b>	23

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE</b>
Figure 3.1: Block diagram of proposed methodology	10
Figure 3.2: Workflow of this project	12
Figure 4.1: User Interface	18

## LIST OF TABLES

<b>TABLES</b>	<b>PAGE</b>
Table 4.1: Correlation between the Attributes	13
Table 4.2: Confusion Matrix for Support Vector Machine	14
Table 4.3: Precision Recall for Support Vector Machine	15
Table 4.4: Confusion Matrix for Decision Tree	16
Table 4.5: Precision Recall for Decision Tree	17

# CHAPTER 1

## Introduction

### 1.1 Introduction

At the end of 2019, in Wuhan, China a novel corona virus designated as 2019-nCov have emerged. 2019-nCoV is the third coronavirus to develop within the human populace within the past two decades and a development that has put worldwide open wellbeing educate on tall caution. There were moreover extreme intense respiratory disorder coronavirus (SARSCoV) episode in 2002 and the Middle East respiratory disorder coronavirus (MERS-CoV) episode in 2012.

After discovery of the causative agent, China reacted rapidly by advising the World Wellbeing organization (WHO) of the flare-up and sharing grouping data with the universal community. The illness related with 2019-nCoV shows up to be moderately mellow as compared with SARS and MERS. Concurring to world health organization (WHO), Corona infections make up a huge family of infections that can contaminate feathered creatures and well evolved creatures, counting people.

Corona viruses have a place to the subfamily Crown virinae within the family Corona viridae. Diverse sorts of human coronaviruses change in how serious the coming about malady gets to be, and how distant they can spread. Specialists right now recognize seven sorts of coronavirus that can contaminate people. Among them common sorts are- 229E (alpha coronavirus), NL63 (alpha coronavirus), OC43 (beta coronavirus), HKU1 (beta coronavirus). SARS-CoV, the infection dependable for serious intense respiratory disorder (SARS). In 2019, a perilous unused strain called SARSCoV-2 begun circulating, causing the infection COVID-19 [1].

Constrained inquire about is accessible on how HCoV spreads from one individual to another. Be that as it may analysts accept that the infections transmit through liquids within the respiratory framework, such as mucus. Without covering the mouth coughing and wheezing can scatter beads into the air. With an individual who has the infection touching and shaking hand can pass the infection between people. Making contact with a

surface or object that has the infection and after that touching the nose, eyes, or mouth may spread the infection as well. Three group of people are highly in risk being effect by this virus. Youthful children, Individuals matured 65 years or more seasoned, Ladies who are pregnant are most helpless to this infection. The World Health Organization (WHO) have declared a public health emergency and addressed corona virus as Pandemic. The primary individuals with COVID-19 had joins to creature and fish advertise. This fact recommended that creatures at first transmitted the infection to people.

## **1.2 Motivation**

In our country, covid situation is much more worse than it looks. So many people having corona symptoms doesn't show any intent or interest to test whether they are effected by corona virus or not. As the Covid situation is getting worse day by day and there is very little amount of test being done every day, I think if I can build such an system where user can give information about his/her symptoms and some actions the system will show the possibility of having corona virus. To complete the test in hospital someone needs to stand in a long queue without the assurance that he/she can give sample for test, as the number of test conducting every day is very little compared to our population. In the queue most of the time there is no social distance maintaining and it is painful for those persons who are suffering from corona symptoms like breathing problem or chest pain standing in the long queue.

## **1.3 Rationale of the Study**

This project has long term impact. For example the earlier works related to similar topic has been theoritical mostly like what is corona virus and how does it spread? None of them was used in real life device, just inputing symptoms and actions. That's the main gap between the study and the studies have been accomplished before. That special problem will be focused throughout the project. And most importantly the result of the study will be used in the real life implementation such devices for people to easily classify whether ther are effected by corona virus or not.

## **1.4 Research Question**

I have created a few inquire about questions which has been stated the issues of my project focuses on and outlined the tasks that have to be completed. The reseach questions are as follows:

- (a) Does the data only associated with corona virus or other disease as well?
- (b) Which classify techniques will be used accomplishing the satisfied and optimal classification accuracy and why?

## **1.5 Expected Output**

There are not so many work regarding this topic. Very few people work with this topic and most of them are theoretical work like what is corona virus, how does it spread and what are the types of corona virus? In this project, a few data has been utilized to prepare the model and a few data is utilized to test the model, finally as a result I have the accuracy for two different algorithm. This research work will accomplish the both, which mean accuracy along with the possibility of having Corona virus. It focuses on whether someone having corona virus or not by analyzing the symptoms and actions.

## **1.6 Report layout**

### **Chapter 1: Introduction**

The very fundamental of my project will be focused in this section. As per project, what is Corona Virus actually and how I got the idea of this project is the main point here. Besides some questions related to my project will be come in front. And the most importantly about the expected output of the project will be discussed in nutshell.

## **Chapter 2: Background**

From where I have got the idea about project will be discussed here as well, but in briefly. For that the literature review of some related works will be come in front. For example what they have done, what have not and what couldn't able to implement. And from that what is the scope of my project in the same area yet will be mainly focused in the background section. Moreover what kind of challenges will have to face to get it done will be discussed here as well.

## **Chapter 3: Research Methodology**

As per the title of this section the main goal of this part is form a research methodology. Most specifically how the project will be get done through which procedure. That's the main area of discussion here. Since the main component of the project is data, so the main focus will be on it. For example how to collect data, which and how much algorithm will be used here to analyze the data and will be used in particular case like pre-processing, testing and so on. A short description of every single algorithm which will be used will be given here. And finally the whole stuffs will be shown very easily with the help of some diagram and with a paper segment of the main methodology.

## **Chapter 4: Experimental Result and Discussion**

Here comes the most import part of the project where all the mathematical terms will be discussed briefly. The experimental result and descriptive will be given in this section which will focus on the mathematical terms of this project and end of the section a through simple short way the whole section will be summarized.

## **Chapter 5: Summary, Conclusion, Recommendation and Implication for future research**

In this section the whole project will be summarized very shortly. And the recommendation and the further implication will be given. There will be comparison among the result and the best algorithm with higher accuracy will be chosen. More specifically what is the ultimate result of the project and how the result can be used in the same field for more advanced in the field of machine Learning.



## **CHAPTER 2**

### **Background**

#### **2.1 Introduction**

The main purpose of this project is to classify whether the patient can be affected by Corona Virus in the most appropriate way and get maximum level of accuracy. Maximum type of works related to Corona Virus has been done before but in different way and most of them were done theoretically. Basically what I have been chosen to accomplish is identifying and observing the symptoms and actions of Corona Virus affected people.

#### **2.2 Related Work**

I am not the very first to work on this topic so far so much research work has been done on this. So many renowned scholars have been working on this for many long.

2019-nCoV shows up to be less pathogenic than MERS-CoV and SARS-CoV but hypertension, diabetes, and heart infection are variables that can disable the resistant framework that can make a patient powerless to sickness. Ponders have appeared that SARS-CoV caused an increment in infection resistance in 2002-2004 epidemic with hereditary changes, return and rise of antigenic variations that were adjusted to the host. The rise of more safe strains complicates and complicates treatment approaches [2].

It appears that Corona Virus is more stable at 4 ° C but very much vulnerable to acidic pH and heat [3].

The number of white blood cells can change in patients with COVID-19. The number of lymphocytes and leukocytes shows up to diminish in the early stages of infection. Death occurred due to high D-dimer levels and severe Lymphocytopenia [4].

Corona virus infections very much spread to the host body like a mouse hepatitis virus. Respiratory, gastrointestinal, and neurological disorders and many more are caused by this virus [5].

Initially the infection was transmitted from animals to people, but presently transmitting quickly from human to human. As the root of the disease is China, but as a result of human-to-human transmission it has reached more than 220 countries of the world [6].

To prevent transmission, suspected patients with corona virus respiratory syndrome infection should be hospitalized in isolation units. They also need proper monitoring 24/7 [7].

In close proximity to the infected person human-to-human transmission occurs. The contaminated individual wheezes and that respiratory droplets, just like the spread of flu, can break down within the verbal or nasal mucosa and lungs of individuals who breathed in the contaminated discuss and that is how transmission of the virus occurs [8].

The COVID-19 infection did not transmit from mother to embryo for pregnant ladies who were in their final trimester of pregnancy. It is imperative to note that, this ponder was conducted in a little number of people [9].

Chest torment, tipsiness, and queasiness were observed in ponder of indications of COVID-19. Loose bowels may be a side effect of with MERS-CoV and SARS-CoV but it is barely seen in COVID-19 [10].

Blood sugar, hypertension and cardiovascular disease are a few side impacts which were found in half of these patients. In addition, patients with anomalous dyspnea may show that these complications are likely to be critical supporters to the passing of COVID-19 patients [11].

In physical structure and pathogenic behavior there is similarity between SARS-CoV-2's and SARS-CoV. To bind to target cells and a cellular protease to prime the S protein CoVs utilize a protein called a spike (S) protein. Between [SARS-CoV-2] spike and SARS-CoV spike around 76%-78% sequence similarities for the whole protein, for the RBD [receptor binding domain] 73%-76%, and for the RBM [receptor binding motif] 50%-53% [12].

IgM and IgG ELISA location units utilizing bat SARS rCoV Rp3 NP were created with no cross-reaction against human coronaviruses but SARSr-CoV.26The methods of ELISA for the assurance of SARSCoV-2 IgG were portrayed some time recently [13].

The most, quickest, and most touchy test for the conclusion of SARS-CoV-2 disease is Nucleic corrosive discovery. For successful location of the primary 25 positive cases of

disease, two settled RT-PCR and two real-time RT-PCR measures have been created in Japan [14].

Antiviral treatment for 2019-nCoV contamination has not been demonstrated to be successful. Patients ought to be treated in isolation. Controlling the source of infection is currently the approach to this disease; utilize of individual security safeguard to decrease the chance of transmission; and early determination, confinement, and steady medications for influenced patients [15].

### **2.3 Research Summary**

My main aim is to find out the optimal solution from different symptoms and actions of Covid patient. I have used data cleaning (Null the row If there is any blank column) technique for data pre-processing. For classification I used Support Vector Machine (SVM) and Decision tree algorithm. Based on accuracy I highly recommend Decision tree algorithm between Support Vector Machine (SVM) and Decision tree for its high accuracy.

### **2.4 Scope of the Problem**

The principal scope of my research is to end out the optimal algorithm in terms of classification for symptoms and actions of Covid patients. The main objective of this project is consummate with the performance analysis between Support Vector Machine (SVM) and Decision tree classifiers.

### **2.5 Challenges**

The most faced problem in any sort of analysis is *Data*. In my case I also faced the same problem. My plan was to use real time data for analysis but I couldn't have managed to talk with the patients or their relatives because I have huge probability of being affected by physical contact with patient and doctor also prohibit me to do such thing. I have been looking for the positive response from "Institute of Epidemiology Disease Control and

Research” for data collection several times but I didn’t get any positive response from their side at first. But when I convinced one doctor of IEDCR to hear my proposal she seems bit convinced and suggest me to write proper proposal and approach officially for the data I need. Luckily a team of few members of IEDCR were working on same topic but in different field, they agreed to provide me a sample dataset and many more important information about patients which helped me to attribute the dataset properly. Finally I have allocated attribute successfully and accomplished my project.

## Chapter 3

### Research Methodology

#### 3.1 Introduction

Developing a methodology is the main task of a research work. Following step by step of the methodology any research work can be accomplished. I have been working with the symptoms. As data I have the binary value of the symptoms and actions for the patients where 0 means No and 1 means Yes.

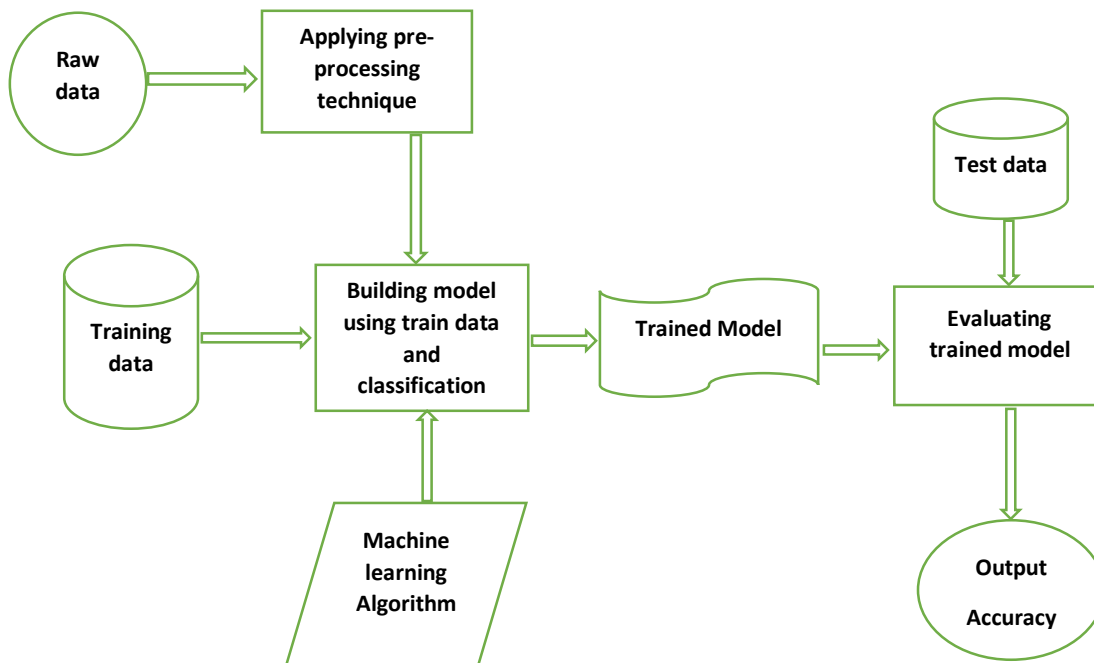


Figure 3.1 Block diagram of proposed methodology

### 3.2 Research Subject and Instrumentation

The main objective of this research is to classify having corona virus or not of the patient. Symptoms are obtained from the calamity of the illness of Covid patients. So, patient having corona virus or not will very much depends on their symptoms of illness.

### 3.3 Data Collection Procedure

The data I have been working with is from “Institute of Epidemiology Disease Control and Research”. First of all I search different online source for proper dataset and also mailed at World Health Organization (WHO) but I failed to get any type of positive response. Then I pay visit to “Institute of Epidemiology Disease Control and Research” for data collection several times but I didn’t get any positive response from their side at first. But when I convinced one doctor of IEDCR to hear my proposal she seems bit convinced and suggest me to write proper proposal and approach officially for the data I need. Luckily a team of few members of IEDCR were working on same topic but in different field, they agreed to provide me a sample dataset and many more important information about patients which helped me to attribute the dataset properly.

### 3.4 Statistical Analysis

**a) Data preprocessing:** Data cleaning is most widely used preprocessing technique in data preprocessing stage. By using this I replace missing values in my dataset for improving accuracy and accomplish better performance.

**Replace missing values:** As I have used ‘pandas’ library in my code for data preprocessing it has 3 basic rules for replacing missing values and those are-

- I. Erase whole row in case there's lost value.
- II. Erase whole row in case there's lost value and also based on some condition for other attributes.
- III. Replace the missing value point with mean value.

As I have to maintain high level of accuracy and proper dataset size so I used the 1st technique which is delete entire row if there is missing value.

**b) Classification using Support Vector Machine and Decision Tree:** Corona Virus classification using machine learning requires to choose an accurate classifier for proper classification and improving accuracy. Both the classifier used in machine learning widely.

**Support Vector Machine Classifier:** A Support Vector Machine could be a classification calculation to relegate information to different classes. It includes recognizing hyperplanes which isolate information into classes.

**Decision Tree:** It has a place to the family of supervised calculation. Decision tree makes a difference to form a preparing demonstrate that can utilize to foresee the lesson or esteem of the target variable by learning basic choice rules.

### 3.5 Implementation Requirements:

For the implementation of proposed research plan I have developed a methodology.

I have segmented the whole workflow into three different steps. The entire research work abide by this three steps.

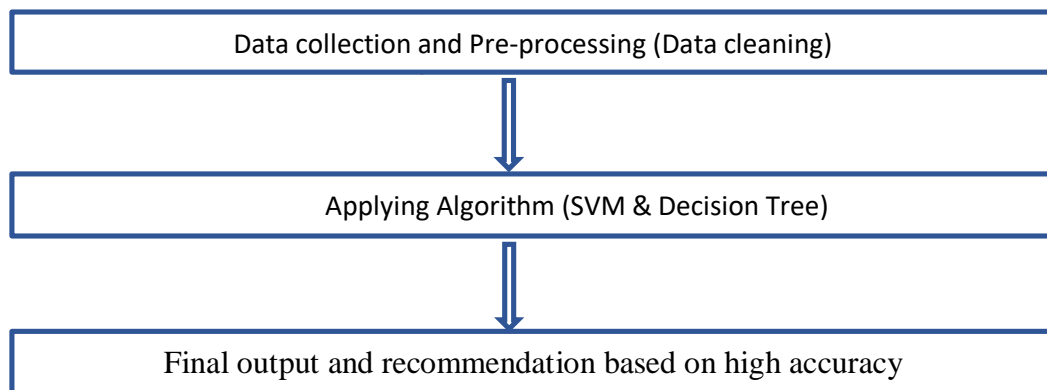


Figure 3.2: Workflow of this project

# CHAPTER 4

## Experimental Result and Discussion

### 4.1 Introduction

In this part, I represent the results of different methodologies and the result of different algorithm which I have applied in the dataset.

### 4.2 Experimental Result

In data preprocessing stage, data cleaning (delete entire row if any blank column) is used.

Support Vector Machine (SVM) and Decision Tree are used here .Support Vector machine give me the accuracy of 85% comparing with the accuracy of Decision Tree algorithms 87%.

### 4.3 Descriptive Analysis

After data preprocessing dataset is absolutely ready to apply classification algorithm.

Here is the correlation between all the 17 symptoms.

Table 4.1: Correlation between the Attributes

	Breathing Problem	Fever	Dry Cough	Sore throat	Running nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	Hyper Tension	Fatigue	Gastrointestinal	Abroad travel	Contact with COVID Patient	Visited Public Exposed Places	Wearing Masks	Class
Breathing Problem	1.000000	0.598280	0.607113	0.658009	0.104998	0.108183	-0.178253	-0.088293	0.002253	0.057033	0.124904	-0.018581	-0.099905	0.420402	0.414909	0.129817	NaN	0.569521
Fever	0.598280	1.000000	0.357112	0.595499	0.134742	0.098516	-0.110898	-0.057721	-0.087921	0.091322	0.137155	-0.088751	-0.069388	0.351654	0.388348	0.083811	NaN	0.450670
Dry Cough	0.607113	0.357112	1.000000	0.491033	-0.042081	0.100519	-0.024851	-0.028877	0.009091	0.018312	0.079005	-0.088200	0.023516	0.383775	0.314248	0.112855	NaN	0.549432
Sore throat	0.658009	0.595499	0.491033	1.000000	0.114410	0.151754	-0.105488	-0.080702	-0.070706	0.070350	0.118025	0.007050	0.041949	0.405981	0.391990	0.137374	NaN	0.574432
Running nose	0.104998	0.134742	-0.042081	0.114410	1.000000	-0.033082	-0.049215	0.126581	-0.058308	0.030932	-0.088117	-0.003743	-0.004380	0.177894	0.058102	0.097478	NaN	0.021642
Asthma	0.108183	0.098516	0.100519	0.151754	-0.033082	1.000000	-0.074998	0.074499	0.057081	-0.053291	-0.009693	0.274417	0.128529	0.028159	0.045009	NaN	NaN	0.108167
Chronic Lung Disease	-0.178253	-0.110898	-0.024851	-0.105488	-0.049215	-0.074998	1.000000	-0.114281	-0.043004	0.078785	0.005734	-0.058653	-0.138751	-0.182007	-0.114709	-0.111182	NaN	-0.073209
Headache	-0.088293	-0.057721	-0.028877	-0.080702	0.126581	0.074499	-0.114281	1.000000	0.008888	0.017750	-0.293892	0.030859	0.120546	0.029093	-0.075888	0.044605	NaN	-0.081288
Heart Disease	0.002253	-0.087921	0.009091	-0.070706	-0.058308	0.057081	-0.043004	0.008888	1.000000	-0.033171	0.141254	-0.080800	0.001603	-0.097516	-0.012598	0.188246	NaN	-0.010899
Diabetes	0.057033	0.091322	0.018312	0.070350	0.030932	-0.053291	0.078785	0.017750	-0.033171	1.000000	0.074888	-0.103290	0.103899	0.094098	-0.058947	-0.118288	NaN	0.038803
Hyper Tension	0.124904	0.137155	0.079005	0.118025	-0.088117	-0.009693	0.005734	-0.293892	0.141254	0.074888	1.000000	-0.054702	-0.073829	-0.024185	0.052283	0.041254	NaN	0.113015
Fatigue	-0.018581	-0.088751	-0.088200	0.007050	-0.003743	-0.018387	-0.058653	0.030859	-0.080800	-0.103290	-0.054702	1.000000	0.038458	-0.047454	-0.048720	-0.028935	NaN	-0.028850
astointestinal	-0.099905	-0.069388	0.023516	0.041949	-0.004380	0.274417	-0.138751	0.120546	0.001603	0.103899	-0.073829	0.038458	1.000000	0.133000	-0.016822	-0.082432	NaN	-0.008051
Abroad travel	0.420402	0.351654	0.383775	0.405981	0.177894	0.128529	-0.182007	0.029093	-0.097516	0.094098	-0.024185	-0.047454	0.133000	1.000000	0.289633	0.102259	NaN	0.390595
Contact with COVID Patient	0.414909	0.388348	0.314248	0.391990	0.058102	0.028159	-0.114709	-0.075888	-0.012598	-0.058947	0.052283	-0.048720	-0.016822	0.289633	1.000000	0.139297	NaN	0.339886
Visited Public exposed Places	0.129817	0.063811	0.112855	0.137374	0.097478	0.048009	-0.111182	0.044605	0.188246	-0.118288	0.041254	-0.028935	-0.082432	0.102259	0.139297	1.000000	NaN	0.134130
Wearing Masks	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Class	0.569521	0.450670	0.549432	0.574432	0.021642	0.108167	-0.073209	-0.081288	-0.010899	0.038803	0.113015	-0.028850	-0.008051	0.390595	0.339886	0.134130	NaN	1.000000

The strong relations leveled with 0.5 or above among those symptoms.



### Support Vector Machine:

The accuracy is 85% after applying the Support Vector Machine algorithm as I use 20% of the dataset as test data and the system chose it randomly. The summary of predictions result on a classification algorithm is represented by confusion matrix. By broke down each class and counting values the number of redress and inaccurate expectations are represented by confusion matrix.

True Positive (TP)

Interpretation: Anticipated as positive and genuine esteem is positive moreover.

True Negative (TN)

Interpretation: Anticipated as negative and genuine esteem is negative moreover.

False Positive (FP)

Interpretation: Genuine esteem is negative but anticipated as positive.

False Negative (FN)

Interpretation: Real esteem is positive but anticipated as negative.

Table 4.2: Confusion Matrix for Support Vector Machine

True Positive	False Positive
207	71
False Negative	True Negative
32	413

There is predicted values through column and actual values through the row.

The values are among 20% of the dataset.

True Positive (TP) =207

True Negative (TN) =413

False Positive (FP) =71

False Negative (FN) =32

Precision: How much the model predicted correctly from all the classes.

Recall: How many classes the model predicted correctly from positive classes. Higher percentage of recall is very much appreciable.

F1-score: comparing between two models called f-score. For calculating precision and recall it plays a huge role. Instead of arithmetic mean it uses harmonic mean.

Table 4.3: Precision Recall for Support Vector Machine

	Precision	Recall	F1-Score	Support
Class No (0)	0.87	0.74	0.80	278
Class Yes (1)	0.85	0.93	0.89	445
Average/Total	0.86	0.84	0.85	723

Here,

$$\text{Precision} = \frac{TP}{TP+FP} = 0.86 = 86\%$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} = 0.84 = 84\%$$

$$\text{F1-Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} = 0.85 = 85\%$$

$$\text{Specificity} = \frac{TN}{TN+FP} = 0.85 = 85\%$$

$$\text{False positive rate} = \frac{FP}{FP+TN} = 0.146 = 1.5\%$$

$$\text{Accuracy} = \frac{TP+TN}{N} = 0.85 = 85\%$$

## Decision Tree:

The accuracy is 87% after applying the Decision Tree algorithm as I use 20% of the dataset as test data and the system chose it randomly. The summary of predictions result on a classification algorithm is represented by confusion matrix. By broke down each class and counting values the number of correct and incorrect predictions are represented by confusion matrix.

Table 4.4: Confusion Matrix for Decision Tree

True Positive	False Positive
212	61
False Negative	True Negative
35	415

There is predicted values through column and actual values through the row.

The values are among 20% of the dataset.

True Positive (TP) =212

True Negative (TN) =415

False Positive (FP) =61

False Negative (FN) =35

Table 4.5: Precision Recall for Decision Tree

	Precision	Recall	F1-Score	Support
Class No (0)	0.62	0.87	0.84	273
Class Yes (1)	1.00	0.92	0.84	450
Average/Total	0.81	0.89	0.84	723

Here,

$$\text{Precision} = \frac{TP}{TP+FP} = 0.81 = 81\%$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} = 0.89 = 89\%$$

$$\text{F1-Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} = 0.84 = 84\%$$

$$\text{Specificity} = \frac{TN}{TN+FP} = 0.87 = 87\%$$

$$\text{False positive rate} = \frac{FP}{FP+TN} = 0.128 = 12.8\%$$

$$\text{Accuracy} = \frac{TP+TN}{N} = 0.87 = 87\%.$$

### User Web Interface:

I have built a web application to check if user input his/her symptoms whether the system can predict correctly or not.

For that I used 'Streamlit', which is a very popular framework of python to build web application. Among the two algorithm I used, Decision Tree algorithm because it gives higher accuracy than Support Vector Machine Algorithm. So that, I used decision tree algorithm in this web application and the dataset as my database.

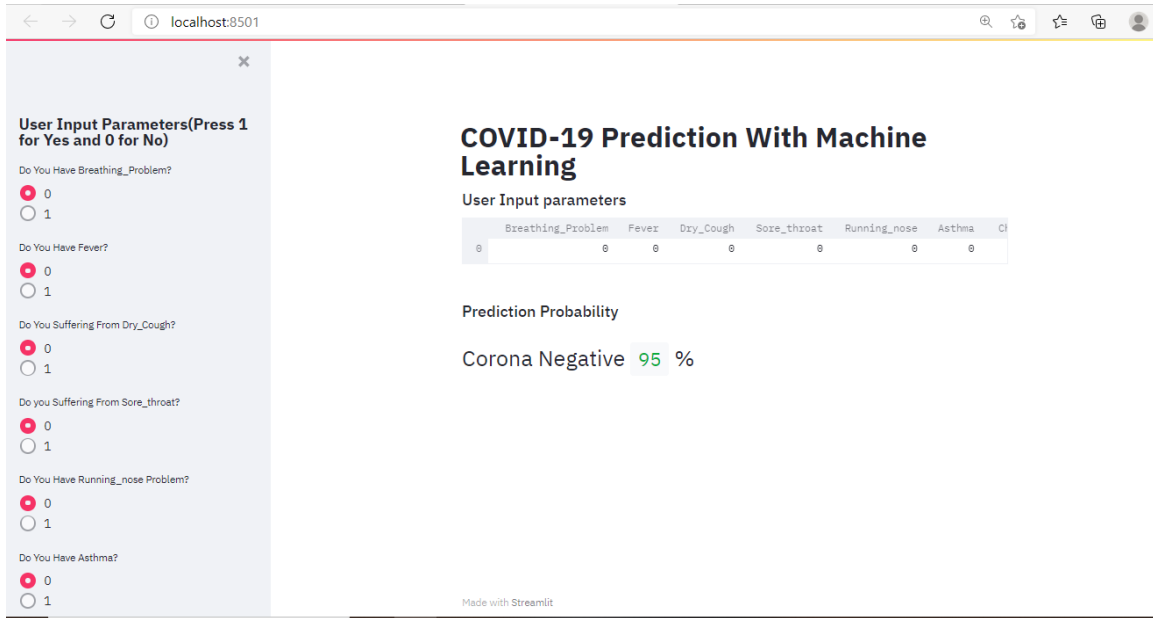


Figure 4.1: User Interface

In my web application, there is user input option in the left side as radio button where user can input his/her symptoms by clicking on those radio button. User have to press 1 if their answer is Yes and press '0' if answer is no. In the right side there is real time user input parameter where user can see what he/she has input. The system compare users input with the dataset and gives the probability of being corona positive or negative as output.

#### 4.4 Summary

After applying two classification Algorithm in processed dataset Support Vector machine provides 85% accuracy where the accuracy of Decision Tree algorithm is 87%. That means Decision Tree algorithm got highest classification accuracy.

## Chapter 5

### Impact on Society, Environment and Sustainability

#### 5.1 Impact on Society:

My project has a huge impact on society. Educated people who are well aware of tech related mechanism they can easily use my system to have an idea whether they have Covid or not. Less educated people who are not well aware of tech mechanism if they can be brought under proper training they can also use my system.

#### 5.2 Impact on Environment:

My project has no such impact on environment as my project is related to health issue of human being.

#### 5.3 Ethical Aspects:

As in our society we treat Corona Virus affected patient differently so it is my duty to keeping secret of patient's personal information and in my project I have maintained this thoroughly.

#### 5.4 Sustainability Plan:

As WHO declared that "Corona Virus and its effect will remain more than two years throughout the world", so my project has long term impact. My project is in initial stage but I would like to build more user friendly web interface and also develop an android app in future. As the symptoms are changing, so I need to change in my data when needed.

## Chapter 6

### Conclusion and Future Scope

#### 6.1 Conclusion:

By the grace of almighty, finally the project “Corona Virus Classification Using Machine Learning Technique in the Context of Bangladesh” and documentation has successfully done. I am happy to complete it after the long term of thinking, discussions and implantations. This project is for those who wanted to get tested if they have corona virus or not. Sometimes it is not possible for everyone to went to the health complex and get tested. They can easily get an approximate idea whether they have corona virus or not and can easily take necessary steps in this regard.

There is no such work has done in this regard. Most of the work on this topic are theoretical like what is corona virus? How it spread and what are the precaution to take. My study result can be used in device or android application which will bring a significant change in the health sector.

#### 6.2 Future Scope:

- Web interface might be more user friendly.
- Using Bengali language in the web application might be more appreciated to common people.
- As symptoms are changing, dataset needs to be updated regularly.
- Check whether Patient’s RNA sequence indicating any change or not.

## References

- [1] S. S. Unhale, Q. B. Ansar, S. Sanap, S. Thakhre, S. Wadatkar, R. Bairagi, P. S. Sagrule and P. D. K. R. Biyani, "A REVIEW ON CORONA VIRUS (COVID-19)," *World Journal of Pharmaceutical and Life Sciences*, vol. 6, pp. 109-115, 2020.
- [2] J. Chen, "Pathogenicity and transmissibility of 2019-nCoV—A quick overview and comparison with other emerging viruses, *Microbes and infection*," Medical imaging consultants Inc., USA, 2020.
- [3] C. L, Y. Y and W. L, "Coronavirus Disease 2019: Coronaviruses and Blood Safety," *Transfusion Medicine Reviews*, 2020.
- [4] D. S and P. H-J, "Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China," *Journal of clinical medicine*, vol. 9(2), p. 575, 2020.
- [5] T. Ksiazek , D. Erdman , C. Goldsmith , S. Zaki, T. Peret and S. Emery , "A novel coronavirus associated with severe acute respiratory syndrome," *The New England journal of medicine*, vol. 348(20), pp. 1953-1966, 2003.
- [6] R. Ralph, J. Lew, T. Zeng, M. Francis, B. Xue and M. Roux, "2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission,travel-related cases, and vaccine readiness," *Journal of infection in developing countries*, vol. 14(1), pp. 3-17, 2020.
- [7] A. Bleibtreu, S. Jaureguiberry, N. Houhou, D. Boutolleau, H. Guillot and D. Vallois, " Clinical management of respiratory syndrome in patients hospitalized for suspected Middle East respiratory syndrome coronavirus infection," *BMC Infectious Diseases*, vol. 18(1), p. 331, 2018.
- [8] S. Riza, A. Erdogan, M. AgaogluPelina , Y. Dineri, C. Yusuf and E. SenelMahmut , "Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature," *Eurasian Journal of medicine and oncology*, vol. 4(1), pp. 1-17, 2020.
- [9] H. Chen, J. Guo, C. Wang, F. Luo, X. Yu and W. Zhang, "Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women," *The Lancet*, vol. 6736(20), p. 3, 2020.
- [10] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong and Y. Han, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *The Lancet*, vol. 395, pp. 507-513, 2020.
- [11] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao and Y. Hu, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, pp. 497-506, 2020.
- [12] Y. Wan, J. Shang, G. Rachel, R. S. Baric and F. Li, "Receptor recognition by a novel coronavirus from Wuhan: An analysis based on decade long structural studies of SARS," *Journal of Virology*, 2020.
- [13] Z. W, X.-L. Yang, W. Y and Y. B, "Molecular and serological investigation of 2019-nCoV infected patients: Implication of multiple shedding routes," *PubMed*, 2020.



- [14] S. K, N. N, K. H, T. L, S. S, H. Katoh, M. Sakata and Y. Nakatsu , "Development of genetic diagnostic methods for novel coronavirus 2019 (nCoV-2019) in Japan," PubMed, 2020.
- [15] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu and J. Xhang, "Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China," Journal of American Medical Association, 2020.

## APPENDICES

Source Code:

<https://drive.google.com/file/d/1gcmHFvOzUWLbVafkbc2lCErpuFJVvps/view?usp=sharing>

## Corona Virus

### ORIGINALITY REPORT

<b>23%</b>	<b>20%</b>	<b>12%</b>	<b>18%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

### PRIMARY SOURCES

<b>1</b>	<b>www.goums.ac.ir</b> Internet Source	<b>4%</b>
<b>2</b>	<b>Submitted to Brickfields Asia College</b> Student Paper	<b>2%</b>
<b>3</b>	<b>Submitted to Daffodil International University</b> Student Paper	<b>2%</b>
<b>4</b>	<b>article.sciencepublishinggroup.com</b> Internet Source	<b>1%</b>
<b>5</b>	<b>www.hindawi.com</b> Internet Source	<b>1%</b>
<b>6</b>	<b>ijaep.com</b> Internet Source	<b>1%</b>
<b>7</b>	<b>dspace.daffodilvarsity.edu.bd:8080</b> Internet Source	<b>1%</b>
<b>8</b>	<b>biomedgrid.com</b> Internet Source	<b>1%</b>
<b>9</b>	<b>Dayi Zhang, Haibo Ling, Xia Huang, Jing Li et al. "Potential spreading risks and disinfection challenges of medical wastewater by the</b>	<b>1%</b>