A Project/Dissertation Report

on

COVID 19 Factors Analysis

Submitted in partial fulfillment of the
requirement for the award of the degree of

Bachelor of Computer Applications

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Mr. Abhay Kumar
Assistant Professor

Submitted By

Palak Mishra19SCSE1010335
Abhishek Nigam 19SCSE1010098

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
MONTH, YEAR

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"COVID 19 FACTOR ANALYSIS"** in partial fulfillment of the requirements for the award of the **Bachelor of Technology, CSE** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Name… Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

<div align="right">

Palak Mishra 19SCSE1010335
Abhishek Nigam 19SCSE1010098

</div>

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

<div align="right">

MR. ABHAY KUMAR

Assistant Professor

</div>

CERTIFICATE

The Final Project/ Dissertation Viva-Voce examination of Palak Mishra (19SCSE1010335) and Abhishek Nigam (19SCSE1010098) has been held on _____ and his/her work is recommended for the award of B.Tech. CSE.

Signature of Examiner(s)   **Signature of Supervisor(s)**

Signature of Project Coordinator   **Signature of Dean**

Date: November, 2013

Place: Greater Noida

Abstract

COVID-19 outbreaks only affect the lives of people, they result in a negative impact on the economy of the country. On Jan. 30, 2020, it was declared as a health emergency for the entire globe by the World Health Organization (WHO). By Apr. 28, 2020, more than 3 million people were infected by this virus and there was no vaccine to prevent. The WHO released certain guidelines for safety, but they were only precautionary measures. The use of information technology with a focus on fields such as data Science and machine learning can help in the fight against this pandemic. It is important to have early warning methods through which one can forecast how much the disease will affect society, on the basis of which the government can take necessary actions without affecting its economy. In this chapter, we include methods for forecasting future cases based on existing data. Machine learning approaches are used and two solutions, one for predicting the chance of being infected and other for forecasting the number of positive cases, are discussed. A trial was done for different algorithms, and the algorithm that gave results with the best accuracy are covered in the chapter. The chapter discusses autoregressive integrated moving average time series for forecasting confirmed cases for various states of India. Two classifiers, random forest and extra tree classifiers, were selected; both have an accuracy of more than 90%. Of the two, the extra tree classifier has 93.62% accuracy. These results can be used to

take corrective measures by different governmental bodies. The availability of

techniques for forecasting infectious disease can make it easier to fight

COVID-19.

Table of Contents

List of Table

S.No.	Caption	Page No.
1		

List of Figures

S.No.	Title	Page No.
1		
2		
3		
4		

Acronyms

B.Tech.	Bachelor of Technology
M.Tech.	Master of Technology
BCA	Bachelor of Computer Applications
MCA	Master of Computer Applications
B.Sc. (CS)	Bachelor of Science in Computer Science
M.Sc. (CS)	Master of Science in Computer Science
SCSE	School of Computing Science and Engineering

CHAPTER 1: INTRODUCTION

In this era of automation, artificial intelligence and data science have important role in the health care industry. These technologies are so well-connected that medical professionals can easily manage their roles and patient care. All health care organizations work hard to develop an automated system that can be used to accept the challenges faced in health care. Scientists are working on machine learning (ML) to develop smart solutions to diagnose and treat disease. ML is capable of detecting disease and virus infections more accurately so that patients' disease can be diagnosed at an early stage, the dangerous stages of diseases can be avoided, and there can be fewer patients. In the same manner, ML can be used to automate the task of predicting COVID-19 infection and help forecast future infection tallies of COVID-19. In this chapter, we include methods for forecasting future cases based on existing data. ML approaches are used and two solutions, one for predicting the chances of being infected and other for forecasting the number of positive cases, are discussed. A trial was done for different algorithms, and the algorithm that gave the results with the best accuracy is covered in the chapter. The chapter discusses autoregressive integrated moving average (ARIMA) time series for forecasting confirmed cases for various states in India. Two classifiers, random forest and extra tree classifier (ETC), are selected; both have an

accuracy of more than 90%. Of the two, ETC has 93.62% accuracy. These results can be used to take corrective measures by different government bodies. The availability of techniques for forecasting infectious disease can make it easier to fight against infectious disease such as COVID-19.

.

INTRODUCTION TO MACHINE LEARNING MODEL -

MACHINE LEARNING: - According to Arthur Samuel (1959), ML is the field of study that gives computers the ability to learn without being explicitly programmed. Thus, we can define ML as the field of computer science in which machines can be designed that can program themselves

The process of learning is simply learning from experience or observations from previous work, such as examples, or instruction, to look for patterns in data and with the help of examples, provided the system can make better decisions. The basic aim of ML is to make computers learn automatically with no human intervention and to adjust perform actions accordingly ast data are used to train the model, and then this trained model is used to test new data and then for prediction. The trained ML model's performance is evaluated using some portion of available past data (which is not present during training). This is usually referred as the validation process.

In this process, the ML model is evaluated for its performance measure, such as accuracy. Accuracy describes the ML model's performance over unseen data in terms of the ratio of the number of correctly predicted features and total available features to be predicted.
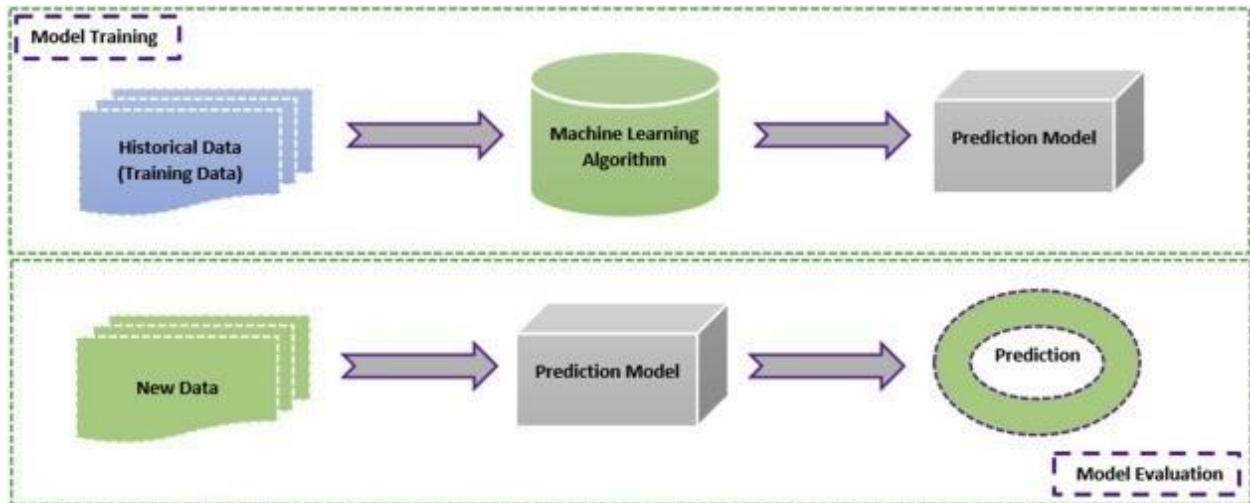
How many people have COVID-19, and what will the pandemic look like in the future? Forecasting models can help predict trends such as infection or mortality rates. But these models rely on data that is collected by different jurisdictions and reported under different standards. This makes it harder to compare data across places and over time. Knowing about these data limitations can improve analysis. For example, since there are challenges tracking COVID-related deaths, COVID-related mortality might be better measured by comparing the number of deaths from all causes during the pandemic with the expected number of deaths for the same period.

Data on Higher-Than-Expected Weekly Mortality for 2020, as of July
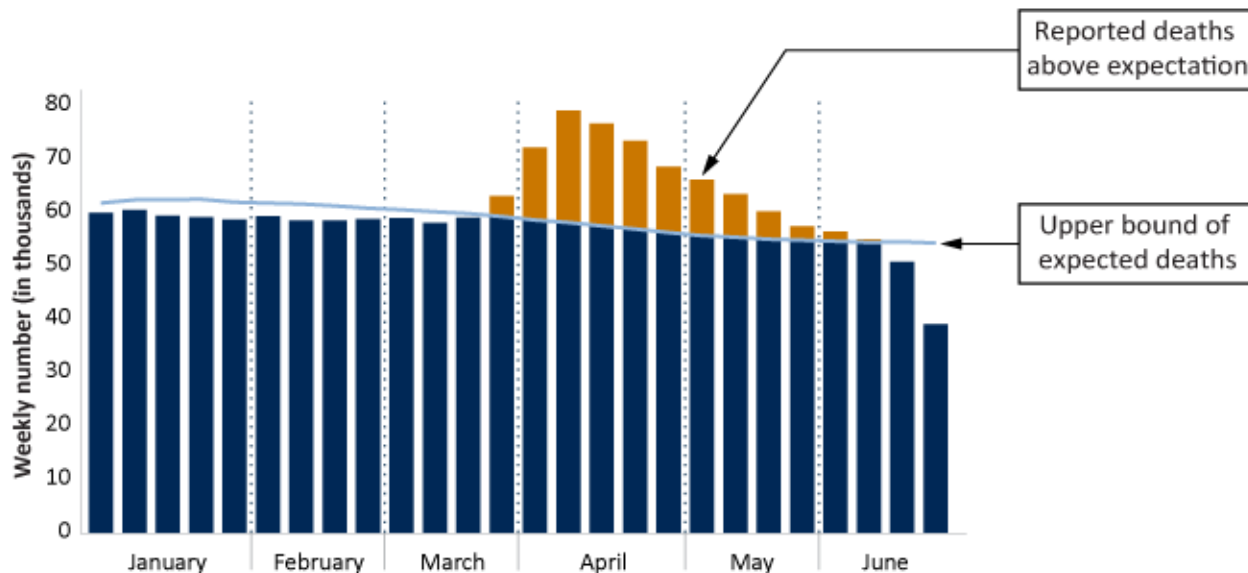
For figure notes, see figure 7 in the report.

Source: GAO analysis of Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS) data. | GAO-20-635SP

Use of machine learning in COVID-

ML is used in various fields, including medicine to predict disease and forecast its outcome. In medicine, the right diagnosis and the right time are the keys to successful treatment. If the treatment has a high error rate, it may cause several deaths. Therefore, researchers have started using artificial intelligence applications for medical treatment. The task is complicated because the researchers have to choose the right tool: it is a matter of life or death

For this task, ML achieved a milestone in the field of health care. ML techniques are used to interpret and analyze large datasets and predict their output. These ML tools were used to identify the symptoms of disease and classify samples into treatment groups. ML helps hospitals to maintain administrative processes and treat infectious disease

ML techniques were previously used to treat cancer, pneumonia, diabetes, Parkinson disease, arthritis, neuromuscular disorders, and many more diseases; they give more than 90% accurate results in prediction and forecasting

The pandemic disease known as COVID-19 is a deadly virus that has cost the lives of many people all over the world. There is no treatment for this virus. ML techniques have been used to predict whether patients are infected by the virus based on symptoms defined by WHO and CDC ML is also used to diagnose the

disease based on x-ray images. For instance, chest images of patients can be used

to detect whether a patient is infected with COVID-19

Moreover, social distancing can be monitored by ML; with the help of this

approach, we can keep ourselves safe from COVID-19

The covid19.analytics package is an open-source tool, which its main implementation and API is the R package. In addition to this, the package has a few more adds-on:

- a central GitHub repository, https://github.com/mponce0/covid19.analytics where the latest development version and source code of the package are available. Users can also submit tickets for bugs, suggestions or comments using the "issues" tab.
- a rendered version with live examples and documentation also hosted at GitHub pages, https://mponce0.github.io/covid19.analytics/
- a dashboard for interactive usage of the package with extended capabilities for users without any coding expertise, https://covid19analytics.scinet.utoronto.ca
- The dashboard can also be deployed locally using the covid19Explorer () function which is part of the covid19.analytics package.
- a *backup* data repository hosted at GitHub, https://github.com/mponce0/covid19analytics.datasets – where replicas of the live datasets are stored for redundancy and robust accessibility's sake.

The following dataset provides facility-level data for hospital utilization aggregated on a weekly basis (Friday to Thursday). These are derived from reports with facility-level granularity across two main sources: (1) HHS TeleTracking, and (2) reporting provided directly to HHS Protect by state/territorial health departments on behalf of their healthcare facilities.
The hospital population includes all hospitals registered with Centers for Medicare & Medicaid Services (CMS) as of June 1, 2020. It includes non-CMS hospitals that have reported since July 15, 2020. It does not include psychiatric, rehabilitation, Indian Health Service (IHS) facilities, U.S. Department of Veterans Affairs (VA) facilities, Defense Health Agency (DHA) facilities, and religious non-medical facilities.

For a given entry, the term "collection_week" signifies the start of the period that is aggregated. For example, a "collection_week" of 2020-11-20 means the average/sum/coverage of the elements captured from that given facility starting and including Friday, November 20, 2020, and ending and including reports for Thursday, November 26, 2020.

Reported elements include an append of either "_coverage", "_sum", or "_avg".

- A "_coverage" append denotes how many times the facility reported that element during that collection week.
- A "_sum" append denotes the sum of the reports provided for that facility for that element during that collection week.
- A "_avg" append is the average of the reports provided for that facility for that element during that collection week.

The file will be updated weekly. No statistical analysis is applied to impute non-response. For averages, calculations are based on the number of values collected for a given hospital in that collection week. Suppression is applied to the file for sums and averages less than four (4). In these cases, the field will be replaced with "-999,999".

This data is preliminary and subject to change as more data become available. Data is available starting on July 31, 2020.

Sometimes, reports for a given facility will be provided to both HHS TeleTracking and HHS Protect. When this occurs, to ensure that there are not duplicate reports, deduplication is applied according to prioritization rules within HHS Protect.

For influenza fields listed in the file, the current HHS guidance marks these fields as optional. As a result, coverage of these elements is varied.

Proposed method for prediction

A symptom-based predictive model was proposed to predict COVID-19 based on symptoms defined by the WHO and CDC

Because there is no proper description of symptoms declared by the WHO, based on some existing symptoms, we defined a model used to predict the disease according to the accuracy given by the model

We created a symptom database in which rules were created and used as input. Then, these data were used as raw data. Then, feature selection took place as part of preprocessing data. The data were divided into training data (80% of data) and test data (20% of data), usually known as the train-test split process. This split is generally done in a stratified or random manner so that population distribution in both groups consists of shuffled data, which leads minimized bias or skewness in the data. Training data were used to train the ML classifier that we used in the model, and test data were used to test that classifier in terms of accuracy received over a predefined unseen portion of the dataset

In our work, the symptoms and patient's class dataset were defined on the basis of symptoms such as fever, cough, and sneezing, whether the patient had traveled to an infected place, age, and whether the patient had a history of disease that could increase the possibly of being infected by the virus

This dataset was then further divided into two sets (training set and testing set) using the test-train split method. The system was trained on the basis of training set data and the accuracy of the ML classifier, and then evaluated over the testing set. Finally, the model was used to predict the probability of infection from the disease using new patient data in terms of positive or negative

Corelation Matrix-Showing pairwise Correlation among Features

1	-0.02102	0.021017	0.018285	-0.0068	-0.00222	0.088825	0.057316	-0.05492	-0.10751	0.107506	-0.03354	0.033544	0.044232	A1
-0.02102	1	-1	0.11257	0.761357	0.018095	0.021696	0.00245	-0.00201	-0.08154	0.081541	-0.1808	0.180797	0.357868	S1
0.021017	-1	1	-0.11257	-0.76136	-0.0181	-0.0217	-0.00245	0.002012	0.081541	-0.08154	0.180797	-0.1808	-0.35787	S2
0.018285	0.11257	-0.11257	1	-0.09358	0.705175	0.261926	0.179949	-0.02426	-0.37884	0.378844	0.186013	-0.18601	0.50586	F1
-0.0068	0.761357	-0.76136	-0.09358	1	-0.02488	-0.04424	-0.02159	0.093096	0.093788	-0.09379	-0.29407	0.294068	0.319788	DC1
-0.00222	0.018095	-0.0181	0.705175	-0.02488	1	0.41441	0.094371	-0.02775	-0.18921	0.189207	-0.00876	0.008764	0.699562	B1
0.088825	0.021696	-0.0217	0.261926	-0.04424	0.41441	1	0.073027	-0.08758	-0.03155	0.031554	0.010201	-0.0102	0.518512	FC
0.057316	0.00245	-0.00245	0.179949	-0.02159	0.094371	0.073027	1	-0.17717	-0.10148	0.101484	0.07626	-0.07626	0.184376	MH
-0.05492	-0.00201	0.002012	-0.02426	0.093096	-0.02775	-0.08758	-0.17717	1	0.183477	-0.18348	-0.12677	0.126769	-0.11489	TH1
-0.10751	-0.08154	0.081541	-0.37884	0.093788	-0.18921	-0.03155	-0.10148	0.183477	1	-1	-0.09379	0.093788	0.076128	LOS1
0.107506	0.081541	-0.08154	0.378844	-0.09379	0.189207	0.031554	0.101484	-0.18348	-1	1	0.093788	-0.09379	-0.07613	LOS2
-0.03354	-0.1808	0.180797	0.186013	-0.29407	-0.00876	0.010201	0.07626	-0.12677	-0.09379	0.093788	1	-1	-0.11089	LOH1
0.033544	0.180797	-0.1808	-0.18601	0.294068	0.008764	-0.0102	-0.07626	0.126769	0.093788	-0.09379	-1	1	0.110887	LOH2
0.044232	0.357868	-0.35787	0.50586	0.319788	0.699562	0.518512	0.184376	-0.11489	0.076128	-0.07613	-0.11089	0.110887	1	Hv-Corona
A1	S1	S2	F1	DC1	B1	FC	MH	TH1	LOS1	LOS2	LOH1	LOH2	Hv-Corona	Features

The correlation coefficient's value near 1 signifies those features participating in correlation are highly correlated to each other; on the other hand, the correlation coefficient's value near 0 signifies those features are less correlated to each other. Generally, correlation could be of two types: positive and negative. A positive correlation states that an increase or decrease in one feature's value results in an increase or decrease in the other feature's value; in contrast, a negative correlation

has a reverse relation between the two features, so an increase in one feature's value results in the decreased value of the other feature.

Rows and columns in the correlation matrix represent each feature's name. Each cell in a table containing the correlation coefficient calculated between features corresponds to the respective row and column of that particular cell.
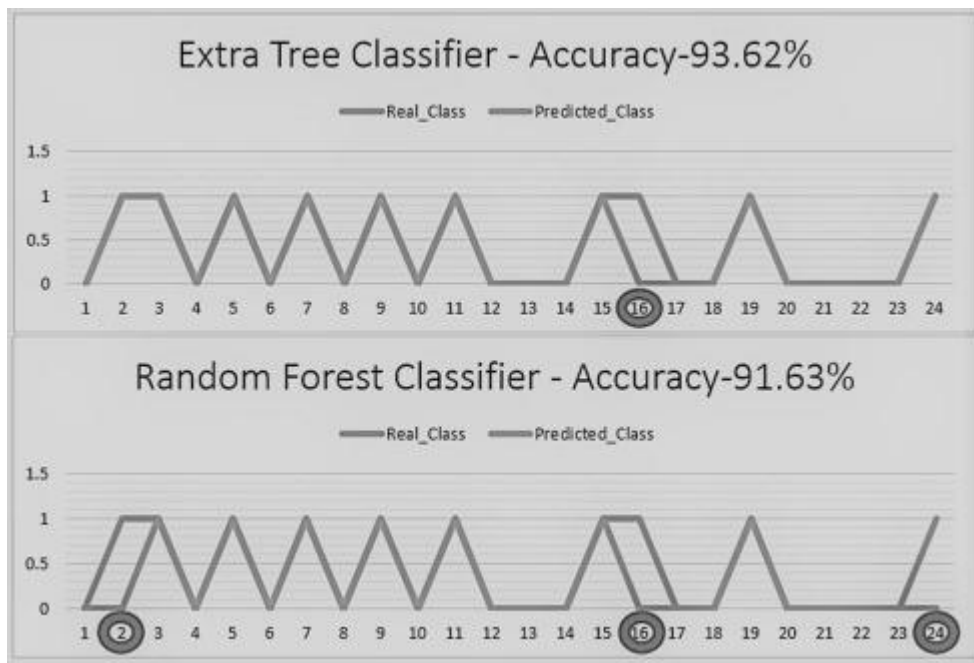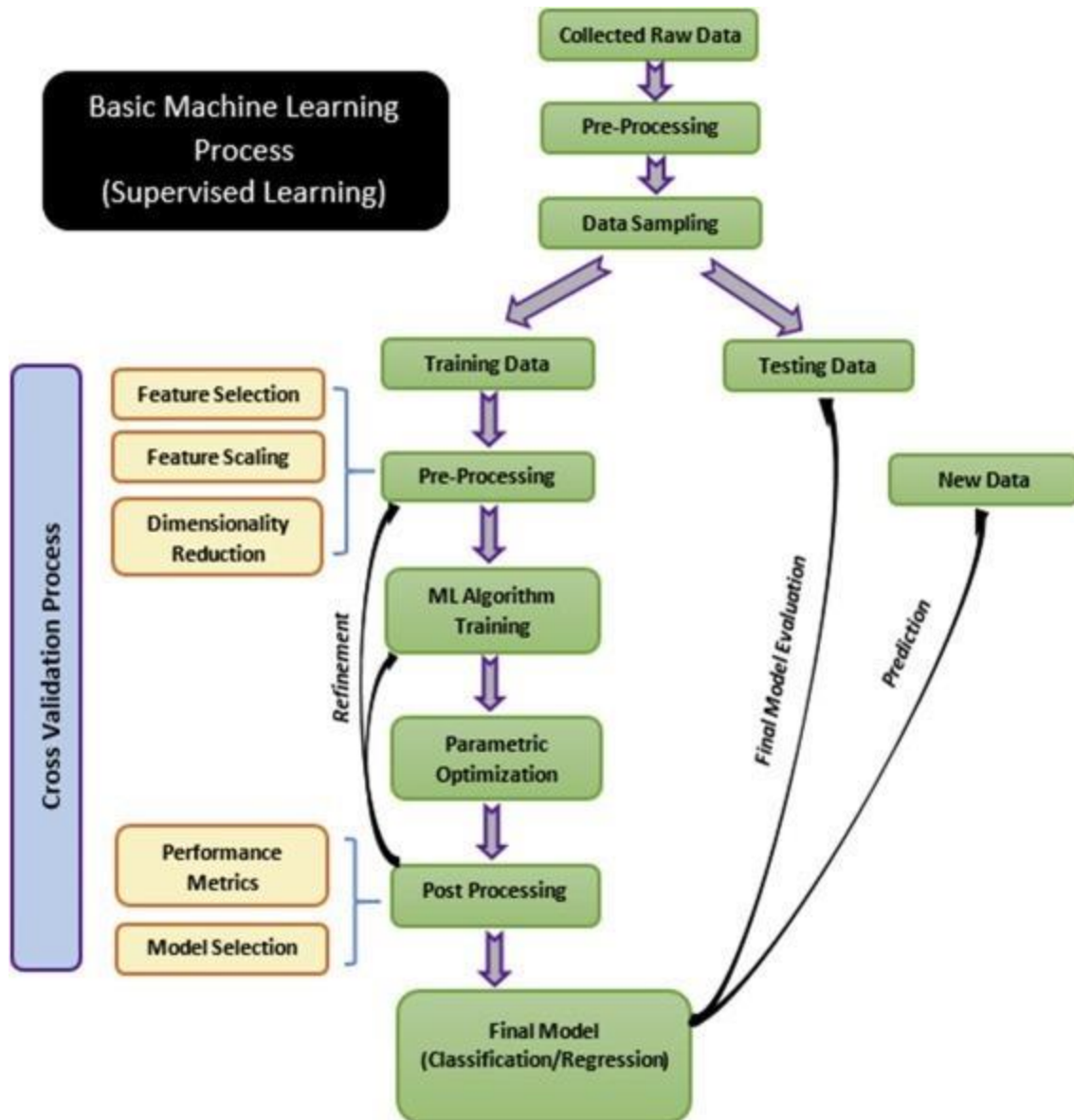
Chapter 2: Literature Survey/Project Design

Different techniques for prediction and forecasting
Various ML techniques are used to predict and forecast future events. Some ML techniques used for prediction are support vector machine, linear regression, logistic regression, naive Bayes, decision trees (random forest and ETC), K-nearest neighbor, and neural networks (multilayer perceptron)
Similarly, some ML techniques used to forecast future events are naive approach, moving average, simple exponential smoothing, Holt's linear trend model, Holt-Winters model, Seasonal Autoregressive Integrated Moving Average Exogenous Model (SARIMAX) and Autoregressive Integrated Moving Average Model (ARIMA). Each technique has unique features and is used differently based on the accuracy results. The model with the best accuracy during the model evaluation process is chosen for prediction or forecasting. In the same way, we identified and used the ETC for the symptom-based prediction of COVID-19 and the ARIMA forecasting model to forecast the number of confirmed cases of COVID-19 in India, because they had the best accuracy results among all classifier and forecasting methods we used when we evaluated model performance.

Basic Machine Learning Process (Supervised Learning)

Collected Raw Data → Pre-Processing → Data Sampling → Training Data / Testing Data

Cross Validation Process

Feature Selection, Feature Scaling, Dimensionality Reduction → Pre-Processing → ML Algorithm Training → Parametric Optimization → Post Processing → Final Model (Classification/Regression)

Performance Metrics, Model Selection

Refinement

Final Model Evaluation, Prediction

New Data

General myths about alcohol and COVID-19

Consuming alcohol destroys the virus that causes COVID-19.

Consuming alcohol will not destroy the virus, and its consumption is likely to increase the health risks if a person becomes infected with the virus. Alcohol (at a

concentration of at least 60% by volume) works as a disinfectant on your skin, but it has no such effect within your system when ingested.

Drinking strong alcohol kills the virus in the inhaled air. Consumption of alcohol will not kill the virus in the inhaled air;
it will not disinfect your mouth and throat; and it will not give you any kind of protection against COVID-19.

Alcohol (beer, wine, distilled spirits or herbal alcohol) stimulates immunity and resistance to the virus.
Alcohol has a deleterious effect on your immune system and will not stimulate immunity and virus resistance.

Alcohol: what to do, and what not to do, during the COVID-19 pandemic

Avoid alcohol altogether so that you do not undermine your own immune system and health and do not risk the health of others.
Stay sober so that you can remain vigilant, act quickly and make decisions with a clear head, for yourself and others in your family and community.
If you drink, keep your drinking to a minimum and avoid getting intoxicated.
Avoid alcohol as a social cue for smoking, and vice versa: people tend to smoke, or smoke more, if they drink alcohol, and smoking is associated with more complicated and dangerous progression of COVID-19.
Remember, too, that indoor smoking is harmful to others in your household and should be avoided.
Make sure that children and young people do not have access to alcohol and do not let them see you consume alcohol – be a role model.

Myth/Fact-
Discuss with children and young people the problems associated with drinking and COVID-19, such as violations of quarantine and physical distancing, which can make the pandemic worse.
Monitor the screen time of your children (including TV), as such media are flooded with alcohol advertising and promotion; they also spread harmful misinformation that may stimulate early initiation and increased consumption of alcohol.
Never mix alcohol with medications, even herbal or over-the-counter remedies, as this could make them less effective, or it might increase their potency to a level where they become toxic and dangerous.
Do not consume alcohol if you take any medication acting on the central nervous system (e.g. pain killers, sleeping tablets, anti-depressants, etc), as alcohol might interfere with your liver function and cause liver failure or other serious problems.

COVID-19 has put health systems under tremendous strain, with countries across the developing world facing their biggest test in living memory. This dashboard keeps track of COVID-19 cases and deaths globally, with a filter on the 72 Gavi-supported countries and Syria. Countries are eligible for Gavi support if their average Gross National Income per capita has been less than or equal to US$ 1,630 over the past three years.
*Pre-filtered on Gavi 72 eligible countries at the start of 4.0 as well as Syria

Alcohol use disorders are characterized by heavy alcohol use and loss of control over alcohol intake. Although they are among the most prevalent mental disorders globally, they are also among the most stigmatized. People with an alcohol use disorder are at greater risk of COVID-19 not only because of the impact of alcohol on their health but also because they are more likely to experience homelessness or incarceration than other members of the population. It is therefore essential, under the current conditions, that people who need help because of their alcohol use get all the support they need.

Seek trusted sources of information, such as WHO, national health authorities and your health professional. For updated information on COVID-19, check the WHO website.3 Always double-check the information you receive. Beware of websites and texts that use the same messages and have the same writing and overall style, as these are likely to be viral messages produced for mass distribution that are intended to mislead. Beware of false and misleading claims, particularly in relation to the effects of alcohol on health and immunity. Such claims should be categorically discounted as a source of health information as there is no evidence that drinking alcohol offers any protection against COVID-19 or has a positive effect on the course and outcomes of any infectious disease. Beware of claims made online that alcohol offers any essential benefits that you really need to have during your period of home isolation or quarantine. Alcohol is in no way a necessary component of your diet and lifestyle.

Early detection of any disease, be it infectious and non-infectious, is critically an important task for early treatment to save more lives [26,27]. Fast diagnosis and screening process helps prevent the spread of pandemic diseases like SARS-CoV-2, cost-effective, and speed up the related diagnosis. The development of an expert system for health care assists in the new order of identification screening and management of SARS-CoV-2 carrier by more cost-effective compared to the traditional method. ML and AI are used to augment the diagnosis and screening process of the identified patient with radio imaging technology akin to Computed Tomography (CT), X-Ray, and Clinical blood sample data.

In this regard, Table 1 shows selective information on diagnosis and screening proposed for the Coronavirus disease. The healthcare expert can use radiology images like X-ray and CT scans as routine tools to augment traditional diagnosis and screening. Unfortunately, the performance of such devices is moderate during the high outburst of the SARS-CoV-2 pandemic. In this regard, studies show the potential of AI and ML tools by suggesting a new model that comes with rapid and valid method SARS-CoV-2 diagnosis using Deep Convolutional Network.

The study shows that diagnosis utilizing an expert system employing AI and ML on 1020 CT images of 108 Covid-19 infected patients along with viral pneumonia of 86 patients, the remarkable performance suggests the use of the convolutional neural network (Resnet-101) as an adjuvant tool for radiologist resulting 86.27%, 83.33% of accuracy and specificity respectively.

ML and AL technology in SARS-CoV-2 Screening.

Publication	ML/AI method	Types of data	No of patients	Validation method	Sample size	Accuracy
Ardakani, A. A *et al*., [28]	Deep Convolutional Neural Network ResNet-101	Clinical, Mamographic	1020, 86	Holdout	1020 CT images of 108 volume of patients with laboratory confirmed Covid-19, 86 CT images of viral and atypical pneumonia patients,	Accuracy: 99.51% Specificity: 99.02%
Ozturk, T. *et al*., [29]	Convolutional Neural Network DarkCovid Net Architecture	Clinical, Mamographic	127, 43 f, 82 m 500, 500	Cross-validation	127 X-ray images with 43 female and 82 male positive cases 500 no-findings and pneumonia cases of 500	Accuracy: 98.08% on Binary classes Accuracy: 87.02% on Multi-classes
Sun, L *et al*., [30]	Support Vector Machine	Clinical, laboratory features, Demographics	336, 220	Holdout	336 infected patients with PCR kit, 26 severe/critical cases and 310 non-serious cases and with another related disease79	Accuracy: 77.5% Specificity: 78.4% AUROC reaches 0.99 training and 0.98 testing dataset

Wu, J. *et al.*, [31]	Random forest Algorithm	Clinical, Demographics	253, 169, 49,24	Cross-validataion	hypertension, 29diabetes, 17 coronary disease and 7 having history of tuberculosis Total of 253 samples from 169 patients suspected with Covid-19 collected from multiple sources. Clinical blood test of 49 patients derived from commercial clinic center. 24 samples infected patient with Covid-19	Accuracy: 95.95% Specificity: 96.95%

Recent studies design an auxiliary tool to increase the accuracy of Covid-19 diagnosis with new model Automatic COVID-19 detection based on deep learning algorithm [29]. The developed model uses raw chest X-ray images of 127 infected patients with 500 no-findings and pneumonia cases of 500 records. With remarkable performance accuracy, binary class of 98.08%, and multi-class with 87.02%. Multi-classes proved the applicability of the expert system to assist radiology in validating in screening process rapidly and accurately.

Furthermore, researchers have found four important medical features combinations of clinical, laboratory features, and demographic information using GHS, CD3 percentage, total protein, and patient age employing Support Vector Machine as the primary feature classification model [30]. The new model is effective and robust in predicting patients in critical/severe conditions, and the empirical results show that a combination of the four-feature results an AUROC of 0.9996 and 0.9757 in training and testing datasets respectively. The survival and the cox-multivariant regression analysis revealed the model's significance towards and auxiliary tools for the healthcare expert.

After evaluating 253 clinical blood samples from Wuhan, researchers found eleven (bilirubin total, creatine kinase isoenzyme, GLU, creatinine, kalium, lactate dehydrogenase, platelet distribution width, calcium, basophil, total protein, and magnesium) key relevant indices which can assist as a discrimination tool of Covid-19 for healthcare expert toward rapid diagnosis [31]. The studies show that 11 relevant indices are extracted after employing the Random Forest algorithm with an overall accuracy of 95.95% and 96.97% specificity respectively. Furthermore, the authors published that the tools were deployed and are available on web-server at http://lishuyan.lzu.edu.cn/COVID2019_2/ to assist healthcare experts.

The above studies give the evidence of an application of the expert system; designing rapid diagnosis was the main objective along with augmentation of accuracy. Prompt and early detection reduce the spread of the disease and reserve more time to the healthcare expert to correspond to the next diagnosis to save more lives, resulting in low-cost medical expenditure. However, majority of the studied paper employed a single classification algorithm on individual data or more.

Therefore, it is suggested to come up with a hybrid classification method applying more potential algorithm on multi-database or hybrid-database consisting of clinical, mammographic, and demographic data, as each type of data has a significant factor that could represent the true identity of the infected patients and deployment of the application in the real world.

ML and AI technology in SARS-Cov-2 contact tracing

If a person diagnoses and is confirmed with Covid-19, the next important step is contact tracing prevention of the wider spread of the disease. According to WHO, the infection spreads from person-to-person primarily through saliva, droplets, or discharges from the nose through contact transmission [32]. To take control on the spread of SARS-Cov-2, contact tracing is an essential public health tool used to break the chain of virus transmission [33]. The process of contact tracing is to identify and manage people who are recently exposed to an infected Covid-19 patient to avoid further spread. Generally, the process identifies the infected person with a follow-up for 14 days since the exposure. If employed thoroughly, this process can break the transmission chain of the current novel coronavirus and suppress the outbreak by giving a higher chance of adequate controls and helping reduce the magnitude of the recent pandemic. In this regard, various infected countries come up with a digital contact tracing process with the mobile application, utilizing different technologies like Bluetooth, Global Positioning System (GPS), Social graph, contact details, network-based API, mobile tracking data, card transaction data, and system physical address. The digital contact tracing process can perform virtually real-time and much faster compared to the non-digital system. All these digital apps are designed to collect individual personal data, which will be analyzed by ML and AI tools to trace a person who is vulnerable to the novel virus due to their recent contacted chain.

As shown in Table 2, articles [34,35] listed various countries competent with such ML and AL-based contact tracing applications. Studies show that over 36 Countries successfully employed digital contact tracing use following centralized, decentralized, or hybrid of both techniques were proposed to lessen the effort and augment the effectiveness of the traditional healthcare diagnosis processes.

Contact tracing application used by Countries.

Sl. No	Country	Contact tracing App	Location tracking	Launch on
1	Australia	COVIDSafe	BlueTrace protocol: Bluetooth	April 14, 2020
2	Austria	Stopp Corona	Bluetooth, Google/Apple	March, 2020
3	Bahrain	BeAware Bahrain	Bluetooth & GSM	March 31, 2020
4	Bulgaria	ViruSafe	GSM	May, 2020
5	China	conjunction with Alipay	GPS, GSM, credit-card-transaction-history	Very little Info
6	Cyprus	CovTracer	GPS, GSM	May, 2020
7	Colombia	CoronApp	GPS	April 12, 2020
8	Czech Republic	eRouška (eFacemask)	BlueTrace protocol: Bluetooth	April 15, 2020
9	Estonia	Estonia's App	Google/Apple, DP-3T, Bluetooth	April, 2020
10	Finland	Ketju	DP-3T, Bluetooth	May, 2020
11	France	StopCovid	Bluetooth	May, 2020
12	Germany	CoronaApp	Bluetooth, Google/Apple	May, 2020
13	Ghana	GH Covid-19 Tracker App	GPS	April 12, 2020
14	Hungary	VírusRadar	Bluetooth	May 13, 2020
15	Iceland	Rakning C-19	GPS	April 2020
16	India	Aarogya Setu	Bluetooth & location-generated social graph	April 2, 2020
17	Iran	Mask.ir	GSM	May, 2020
18	Ireland	HSE Covid-19 App	Bluetooth, Google/Apple	May, 2020
19	Israel	HaMagen	Standard location APIs	March, 2020
20	Italy	Immuni	Bluetooth, Google/Apple	May, 2020

Concerning contact tracing, studies have proven the use of ML and AI in augmentation of contact tracing process against infectious Chronic Wasting disease . After applying Graph theory on infectious animal disease epidemics data, mainly shipment data between each farm, the resultant graph properties generated by the proposed model can be used to exploit to augment contact tracing more efficiently. Moreover, the generated graphs have a potential prediction impact on the number of infections that can take place. However, there are still limitations in addressing the scenario, privacy, control over the data, and even data security breach. Countries are working to overcome the challenges; some countries like Israel "passed an emergency law to use mobile phone data" to tackle the current pandemic. Among the world contact tracing apps, some countries app violated privacy law and reported unsafe so far, they do the job acceptably by supplement the manual tracing process. However, virtually every country has their contact tracing application individually, as the outbreak continues to spread across the world, it becomes a global health emergency. To fight against the Covid-19 as one, one should provide a standard de-facto centralized contact tracing application to trace every human being all around the world. Also, it is reported that some specific query needs to address: "Is it mandatory or voluntary?" "Is the attempt clear or translucent?" "Is information gathering lessened?" "Will collected information be demolished as declared?", "Is the data safe with the host" and "Are there any restrictions or control on utilizing the information?".

ML and AI technology in SARS-CoV-2 prediction and forecasting

Selective information shown in Table 3 indicates the applications of ML and AI in forecasting and predicting the novel pandemic. A new novel model, that forecast and predicting 1-3 to 6 days ahead of total Covid-19 patient of 10 Brazilian states, using stacking-ensemble with support vector regression algorithm on the cumulative positive Covid-19 cases of Brazilian data was proposed, thus augmenting the short-term forecasting process to alert the healthcare expert and the government to tackle the pandemic.

Recent studies suggested a novel model using a supervised multi-layered recursive classifier called XGBoost on clinical and mammographic factor datasets. After applying the model, researchers found out those three significant key features (high-sensitivity C-reactive protein, lymphocyte and lactic dehydrogenase (LDH)) of the 75 features clinical and blood test samples result to be the highest rank of 90% accuracy in predicting and assessing Covid-19 patient into general, severe and mortality rate.

Furthermore, comparatively higher value in single lactic dehydrogenase appears to be a significant factor in classifying most patients in need of intensive medical care, as LDH degree related to various respiratory disorder diseases, namely asthma and bronchitis, and pneumonia.

The forecast model employed decision rule to forecast rapidly and predict infected individuals at the highest risk, authorized patients to be manageable for intensive care, and possibly lessen the transience rate. A Canadian based forecasting model using time-series was developed employing Deep learning algorithm for the long-short-term-memory network, the studies found out a key factor intended for predicting the course with an ending point estimation of the current SARS-CoV-2 epidemic in Canada and all over the globe.

The suggested model forecast ending point of this SARS-CoV-2 outbreak in Canada will be around June 2020. Based on the data collected from John Hopkins University, the prediction was likely to be accurate as newly infected cases have dropped rapidly and proven the applicability of the expert system in predicting and forecasting for the current pandemic outbreak by revealing key significant features. The real-time forecasting model was proposed combining the goodness of the wavelet-based forecasting model and autoregressive integrated moving average-based time-series model.

The model solves the problem by generating short-term forecasts of the SARS-CoV-2 for various countries (India, United Kingdom, Canada, South Korea, and France) to assist healthcare experts and policymakers as a preliminary cautioning module for each target country.

ML and AI applications: prediction and forecasting SARS-CoV-2.

Publication	ML/AI method	Types of data	No of patients	Validation method	Results
Ribeiro, M. H. D. M., *et al.*, [37]	Support Vector Regression and stacking-ensemble	Clinical	40.581	Holdout	Accuracy: Error in range of 0.87%-3.51% one, 1 .02%–5.63% three and 0.95% - 6.90% six day ahead
Yan, L. *et al.*, [38]	XGBoost classifier	Clinical, Blood samples of 75 features	485	Cross-validation	Accuracy: 90%
Chimmula, V.K.R., *et al.*, [39]	Deep Learning using LSTM network	Demographic	John Hopkins University & Canadian Health authority, data containing infected cases upto March 31, 2020	Cross-validation	Ending point of the pandemic outbreak in Canada was predicted on June 2020
Chakraborty, T. and Ghosh, I. [40]	Hybrid Wavelet-autoregressive integrated moving average model and regression tree	Demographic	India: 64 UK: 65 Canada:70 France: 71 South Korea: 76	Cross-validation	Real-time forecast and 10 days ahead, Observed seven key features associated with dead rate.

ML and AI technology in SARS-CoV-2 drugs and vaccination

Since the coronavirus epidemic fury, researchers and healthcare experts around the globe ubiquitously urged to develop a possible choice to tackle the development of drug and vaccine for the SARS-CoV-2 pandemic, and ML/AI technology constitutes to be an enthralling road. Concerning the possibility of drug choice for infected patient's treatment, instant testing on the existing old marketable medicines for novel SARS-CoV-2 carrier in a human being is essential.

Researchers from Taiwan are building a new model to augment the development of a novel drug. After applying the ML and AI technology-based model on two datasets (one using the 3C-like protease constraint and other data-holding records of infected SARS-CoV, SARS-Cov-2, influenza, and human immunodeficiency virus (HIV)) using Deep Neural Network on the eighty old drugs with potential for Covid-19 treatment, the study suggested eight drugs, i.e., vismodegib, gemcitabine, clofazimine, celecoxib, brequinar, conivaptan, bedaquiline and tolcapone are found virtually effective against feline infectious peritonitis coronavirus.

Furthermore, other five drugs like homoharringtonine, salinomycin, boceprevir, tilorone and chloroquine are also found operational during AI experimental environment.

A novel molecule transformer-drug target interaction model was proposed jointly by researchers from the US and Korea to tackle the need for an antiviral drug that can treat the Covid-19 virus. The study compares the accessible virtual screening and molecular docking application called AutoDock Vina with the proposed model employing a deep learning algorithm on 3C-like proteinase of Covid-19 and FDA approved 3,410 existing drugs available in the market.

The result revealed that a popular antiretroviral drug used to treat HIV called Antazanavir (Kd of 94.94 nM) is the best drug for Covid-19 medication, followed by Remdisivir (Kd of 113.13 nM). Furthermore, results revealed that some medications like darunavir, ritonavir, and lopinavir were outlined to tackle viral proteinases. It was also found that various antiviral compounds like Kaletra might be utilized for the medicine of Covid-19 human patients.

A group of researchers from the USA discovered an antiviral drug for treating the Ebola virus. The discovery was first active in the year 2014, starting with ML and AI-based pharmacophore computational analyzing on a limited size of in vitro infected carriers of the Ebola virus.

The study proposed an amodiaquine and chloroquine compound popularly used to treat the malaria virus. Furthermore, after uncovering a decade of drug development based on ML and AI technology, a fusion of computational screening method with docking application and machine learning for choosing supplementary medication to investigate on SARS-CoV-2 was proposed.

Researchers refer to the successful discovery of Ebola, and the Zika virus experience gain belief that the same model could be repeatedly utilized for drug discovery on Covid-19 and future virus pandemic ahead.

The selected review paper adopted various methodologies and technologies addressing the classical method of classification based on statistics to an advanced modern AI and ML algorithm. The use of computational tools, combined with docking application, was found to be more active in predicting the reusability of an existing old drug on Covid-19 medication and dramatically minimize the level of a risk factor in the development of medicine more cost-effective process. During this urgency, the use of ML and AI can augment the drug development process by lessening the time slot on discovering a supplementary treatment and medication for the carrier by drawing a vast probability over security, manageability, and clinical information on the existing drug compound. Issues and challenges found in this area were the limited resource of comprehensive hybrid data and real-life deployment of the application.

During the week 13-19 December, the global number of new cases remained similar to those reported during the previous week; however, the weekly incidence of deaths decreased by 9%. Nonetheless, this corresponds to over 4.1 million new cases and just under 45 000 new deaths. As of 19 December, over 273 million cases and over 5.3 million deaths have been reported globally.

In this edition an update is also provided on SARS-CoV-2 variants of concern (VOCs) and variants of interest (VOIs).

Please note, the next two issues of the Weekly Epidemiological Report (to be published on 28 December and 6 January) will be condensed versions covering only the global and regional epidemiology.

https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

CHAPTER 3: FUNCTIONALITY AND WORKING OF PROJECT

The dataset taken from Kaggle on covid 19 is analyses for various features.

Context

- A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of China's Hubei province
- People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective.
- The virus has shown evidence of human-to-human transmission
- Transmission rate (rate of infection) appeared to escalate in mid-January 2020
- As of 30 January 2020, approximately 8,243 cases have been confirmed

Content

- full_grouped.csv - Day to day country wise no. of cases (Has County/State/Province level data)
- covid*19*clean_complete.csv - Day to day country wise no. of cases (Doesn't have County/State/Province level data)
- country*wise*latest.csv - Latest country level no. of cases
- day_wise.csv - Day wise no. of cases (Doesn't have country level data)
- usa*county*wise.csv - Day to day county level no. of cases

The rapid spread and magnitude of the COVID-19 pandemic have underscored the importance of having quality data, analyses, and models describing the potential trajectory of COVID-19 to help understand the effects of the disease in the U.S. The Centers for Disease Control and Prevention (CDC) is using multiple surveillance systems to collect data on COVID-19 in the U.S. in collaboration with state, local, and academic and other partners. The data from these surveillance systems can be useful for understanding the disease, but decision makers and analysts must understand their limitations in order to interpret them properly. For example, surveillance data on the number of reported COVID-19 cases are incomplete for a number of reasons, and they are an undercount the true number of cases, according to CDC and others.

There are multiple approaches to analyzing COVID-19 data that yield different insights. For example, some approaches can help compare the effects of the disease across population groups. Additional analytical approaches can help to address incomplete and inconsistent reporting of COVID-19 deaths as well.

For example, analysts can examine the number of deaths beyond what would normally be expected in the absence of the pandemic. Examining higher-than-expected deaths from all causes helps to address limitations in the reporting of COVID-19 deaths because the number of total deaths is likely more accurate than the numbers of deaths from specific causes.

The figure below shows actual deaths from the weeks ending January 1 through June 27, 2020, based on data from CDC's National Center for Health Statistics, compared with the expected deaths based on prior years' data. Deaths that exceeded this threshold starting in late March are considered excess deaths that may be related to the COVID-19 pandemic.

Higher-Than-Expected Weekly Mortality for 2020, as of July 14, 2020
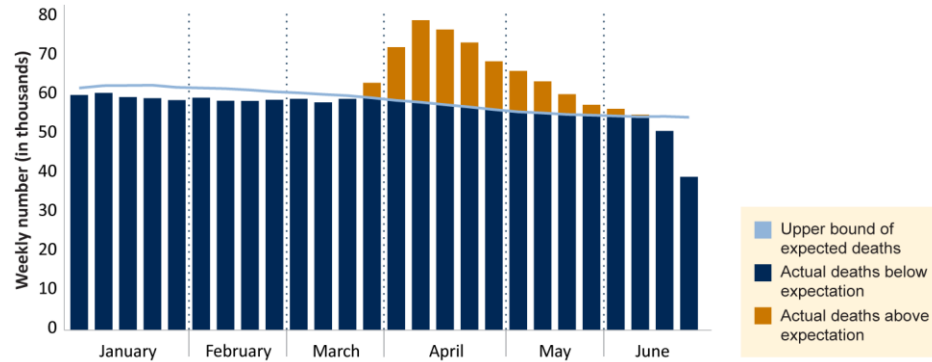
Source: GAO analysis of Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS) data. | GAO-20-635SP

Analysts have used several forecasting models to predict the spread of COVID-19, and understanding these models requires understanding their purpose and limitations. For example, some models attempt to predict the effects of various interventions, whereas other models attempt to forecast the number of cases based on current data. At the beginning of an outbreak, such predictions are less likely to be accurate, but accuracy can improve as the disease becomes better understood.

Why WE Did This Study

The COVID-19 pandemic has resulted in significant loss of life and profoundly disrupted the U.S. economy and society, and the Congress has taken action to support a multifaceted federal response on an unprecedented scale. It is important for decision makers to understand the limitations of COVID-19 data, and the uses and limitations of various methods of analyzing and interpreting those data.
The Coronavirus Aid, Relief, and Economic Security Act (CARES Act) includes a provision for WE to, in general, conduct monitoring and oversight of the authorities and funding provided to address the COVID-19 pandemic and the effect of the pandemic on the health, economy, and public and private institutions of the U.S. This technology assessment examines (1) collection methods and limitations of COVID-19 surveillance data reported by CDC, (2) approaches for analyzing COVID-19 data, and (3) uses and limitations of forecast modeling for understanding of COVID-19.

In conducting this assessment, WE obtained publicly available information from CDC and state health departments, among other sources, and reviewed relevant peer reviewed and preprint (non-peer-reviewed) literature, as well as published technical data on specific models.

```
> df<- read.csv("worldometer_data.csv")
> mean(df$TotalCases)
[1] 91718.5
> getmode <- function(v) {
+     uniqv <- unique(v)
+     uniqv[which.max(tabulate(match(v, uniqv)))]
+ }
>
> getmode(df$Country.Region)
[1] "USA"
> getmode(df$Continent)
[1] "Africa"
> getmode(df$TotalRecovered)
[1] NA
> mode(df$TotalDeaths)
[1] "numeric"
> getmode(df$TotalRecovered)
[1] NA
> mode(df$TotalDeaths)
[1] "numeric"
> getmode(df$TotalRecovered)
[1] NA
> getmode(df$Population)
[1] 331198130
> getmode(df$TotalDeaths)
[1] NA
> getmode(df$NewCases)
[1] NA
> mean(df$TotalRecovered)
[1] NA
```

```
> cummax(df$TotalRecovered)
  [1] 2576668 2576668 2576668 2576668 2576668 2576668 2576668 2576668
  [9] 2576668      NA      NA      NA      NA      NA      NA      NA
 [17]      NA      NA      NA      NA      NA      NA      NA      NA
 [25]      NA      NA      NA      NA      NA      NA      NA      NA
 [33]      NA      NA      NA      NA      NA      NA      NA      NA
 [41]      NA      NA      NA      NA      NA      NA      NA      NA
 [49]      NA      NA      NA      NA      NA      NA      NA      NA
 [57]      NA      NA      NA      NA      NA      NA      NA      NA
 [65]      NA      NA      NA      NA      NA      NA      NA      NA
 [73]      NA      NA      NA      NA      NA      NA      NA      NA
 [81]      NA      NA      NA      NA      NA      NA      NA      NA
 [89]      NA      NA      NA      NA      NA      NA      NA      NA
 [97]      NA      NA      NA      NA      NA      NA      NA      NA
[105]      NA      NA      NA      NA      NA      NA      NA      NA
[113]      NA      NA      NA      NA      NA      NA      NA      NA
[121]      NA      NA      NA      NA      NA      NA      NA      NA
[129]      NA      NA      NA      NA      NA      NA      NA      NA
[137]      NA      NA      NA      NA      NA      NA      NA      NA
[145]      NA      NA      NA      NA      NA      NA      NA      NA
[153]      NA      NA      NA      NA      NA      NA      NA      NA
[161]      NA      NA      NA      NA      NA      NA      NA      NA
[169]      NA      NA      NA      NA      NA      NA      NA      NA
[177]      NA      NA      NA      NA      NA      NA      NA      NA
[185]      NA      NA      NA      NA      NA      NA      NA      NA
[193]      NA      NA      NA      NA      NA      NA      NA      NA
[201]      NA      NA      NA      NA      NA      NA      NA      NA
[209]      NA
> max(df$TotalRecovered)
[1] NA
> cummax(df$TotalDeaths)
  [1] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [10] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [19] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [28] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [37] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [46] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [55] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [64] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [73] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [82] 162804 162804 162804 162804 162804 162804 162804 162804 162804
 [91] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[100] 162804 162804 162804 162804 162804 162804 162804 162804 162804
```

```
[109] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[118] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[127] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[136] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[145] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[154] 162804 162804 162804 162804 162804 162804 162804 162804 162804
[163] 162804 162804 162804 162804 162804 162804     NA     NA     NA
[172]     NA     NA     NA     NA     NA     NA     NA     NA     NA
[181]     NA     NA     NA     NA     NA     NA     NA     NA     NA
[190]     NA     NA     NA     NA     NA     NA     NA     NA     NA
[199]     NA     NA     NA     NA     NA     NA     NA     NA     NA
[208]     NA     NA
> cummax(df$Population)
  [1]  331198130  331198130 1381344997 1381344997 1381344997 1381344997
  [7] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [13] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [19] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [25] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [31] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [37] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [43] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [49] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [55] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [61] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [67] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [73] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [79] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [85] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [91] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
 [97] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[103] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[109] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[115] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[121] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[127] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[133] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[139] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[145] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[151] 1381344997 1381344997 1381344997 1381344997 1381344997 1381344997
[157]     NA     NA     NA     NA     NA     NA
[163]     NA     NA     NA     NA     NA     NA
[169]     NA     NA     NA     NA     NA     NA
[175]     NA     NA     NA     NA     NA     NA
[181]     NA     NA     NA     NA     NA     NA
[187]     NA     NA     NA     NA     NA     NA
[193]     NA     NA     NA     NA     NA     NA
```

```
[199]      NA      NA      NA      NA      NA      NA
[205]      NA      NA      NA      NA      NA
> max(df$Country.Region)
[1] "Zimbabwe"
> max(df$Continent)
[1] "South America"
> hist(df$TotalDeaths)
> dev.off()
null device
          1
> png(file="histogram2.png")
> hist(df$ActiveCases)
> dev.off()
null device
          1
> png(file="histogram3.png")
> hist(df$TotalRecovered)
> dev.off()
null device
          1
> hist(df$TotalTests)
> png(file="histogram4.png")
> hist(df$TotalTests)
> dev.off()
RStudioGD
          2
> png(file="histogram5.png")
> hist(df$ActiveCases)
> dev.off()
RStudioGD
          2
> png(file="boxplot.png")
> boxplot(df$TotalCases)
> dev.off()
RStudioGD
          2
```
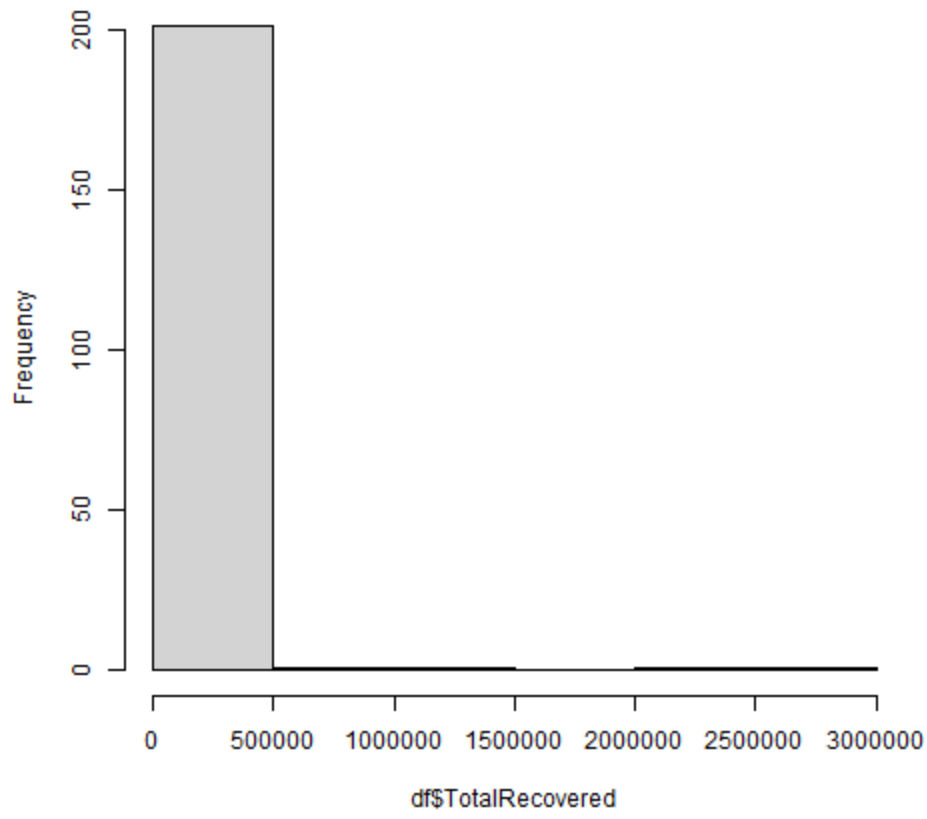
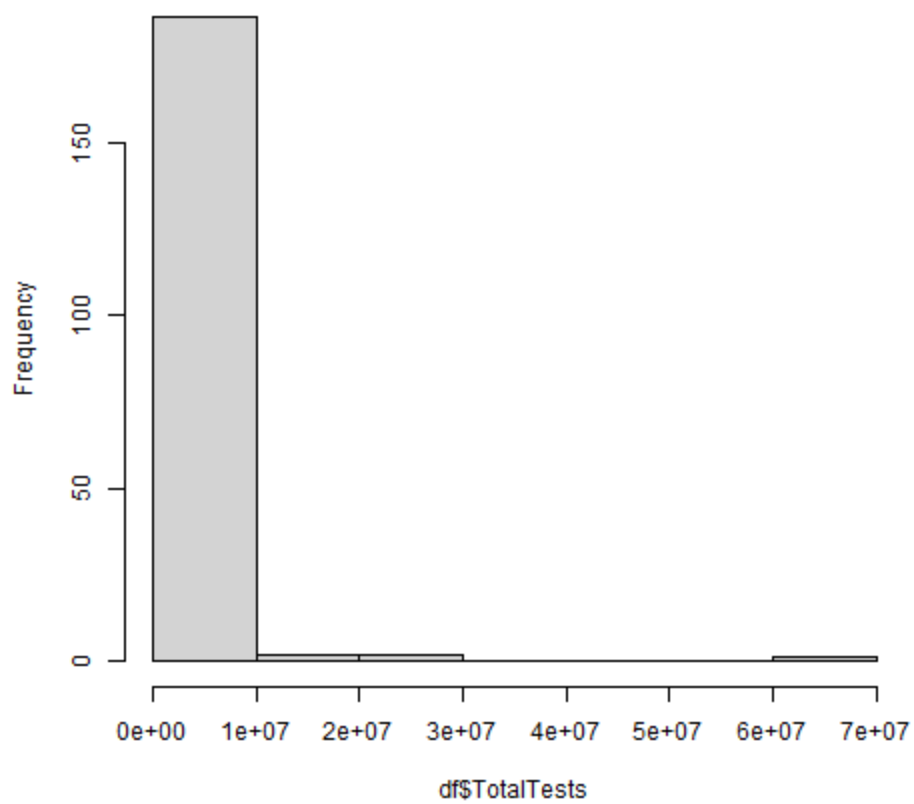Boxplot of total cases
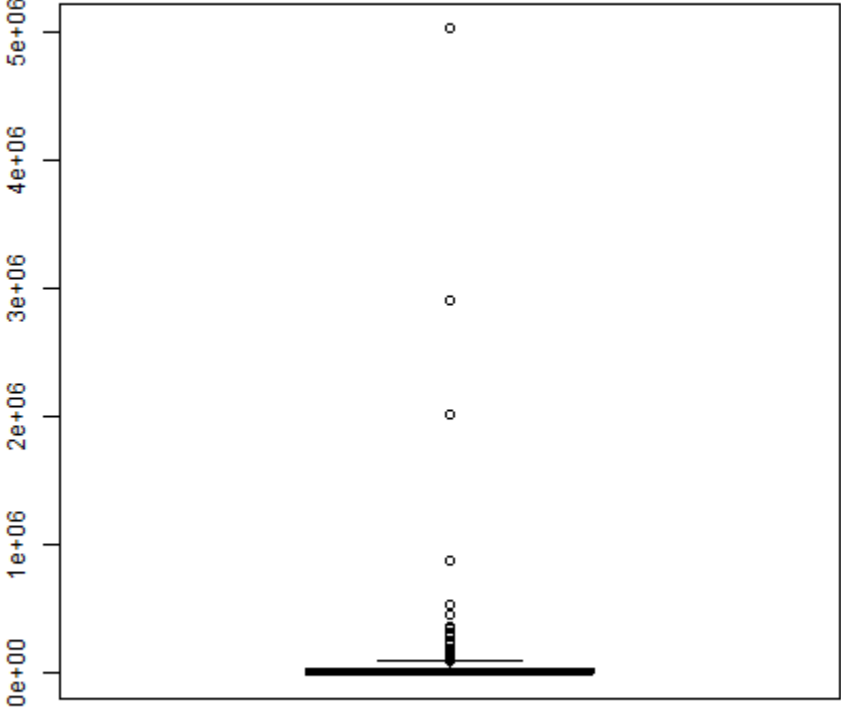
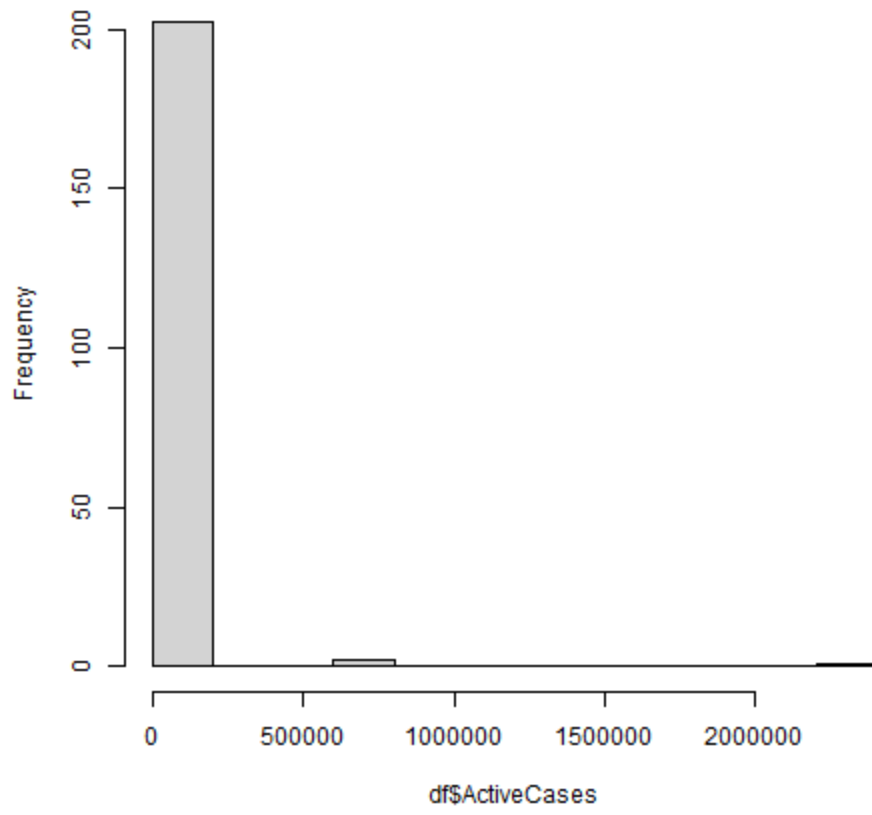**Histogram of df$TotalDeaths**
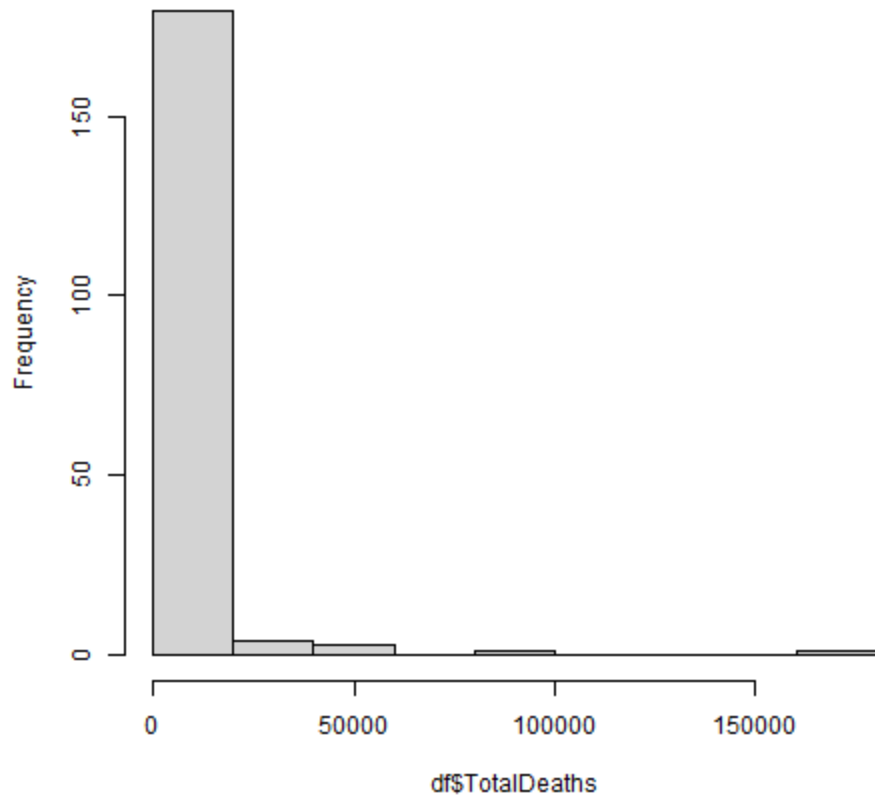
Histogram of df$ActiveCases
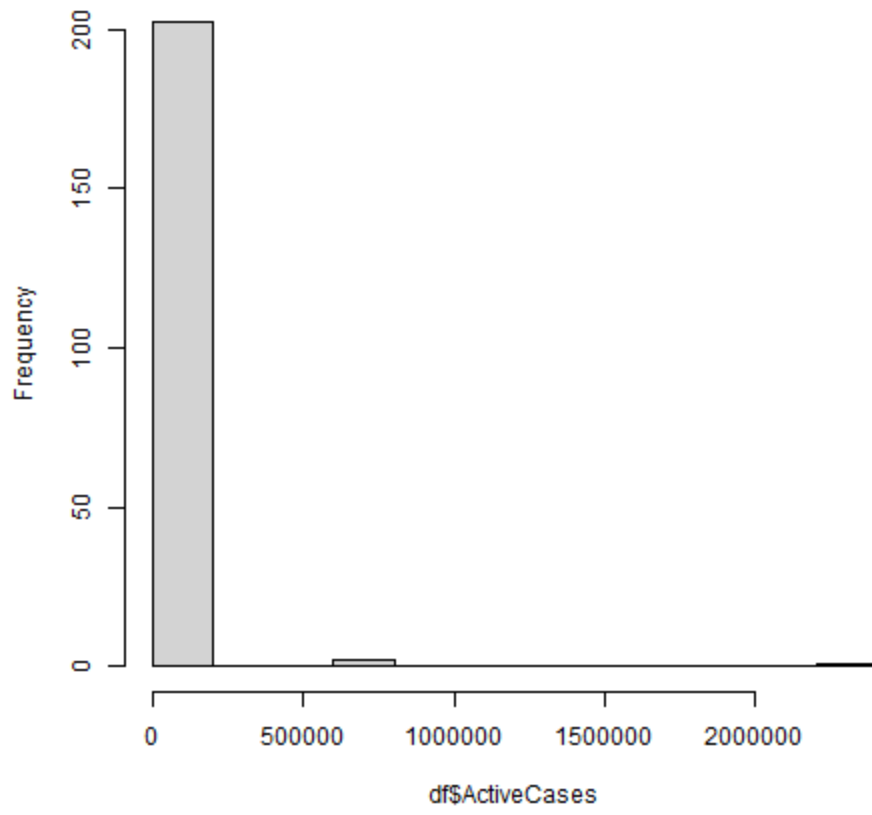
Histogram of df$TotalRecovered

Histogram of df$TotalTests

Histogram of df$ActiveCases

CHAPTER 5: CONCLUSION AND FUTURE SCOPE

The pandemic of COVID-19 has affected the entire globe. It has spread in more than 85 countries as of Apr. 2020. Scientists have made every effort to find solutions to it; according to claims by the United States and India, some vaccines have been made that are being trialed. The use of computers by scientists for early prediction has been widespread. A lot of research is taking place using ML to combat COVID-19. This chapter can be used by different researchers to learn how ML can be employed to forecast not only this situation but also other cases. The chapter specifically used the ARIMA method of time to forecast the stability and growth of COVID-19. Many countries have seen high totals of deaths owing to COVID-19. It is believed that the performance of the model can be improved or the model can give more accurate data if more datasets are available. The model gives results on the basis of data developed by information given by health agencies. Thus, forecasting may not be 100% accurate, but it can surely be used as a corrective measure. For future work further enhancement can be done by combining new factors and algorithms with ARIMA to get more accurate results.

Conclusion and discussion

Since the outbreak of the novel SARS-CoV-2, scientists and medical industries around the globe ubiquitously urged to fight against the pandemic, searching alternative method of rapid screening and prediction process, contact tracing, forecasting, and development of vaccine or drugs with the more accurate and reliable operation. Machine Learning and Artificial Intelligence are such promising methods employed by various healthcare providers. This paper addresses on recent studies that apply such advance technology in augmenting the researchers in multiple angles, addressing the troubles and challenges while using such algorithm in assisting medical expert in real-world problems. This paper also discusses suggestions conveying researchers on AI/ML-based model design, medical experts, and policymakers on few errors encountered in the current situation while tackling the current pandemic. This review shows that the use of modern technology with AI and ML dramatically improves the screening, prediction, contact tracing, forecasting, and drug/vaccine development with extreme reliability. Majority of the paper employed deep learning algorithms and is found to have more potential, robust, and advance among the other learning algorithms. However, the current urgency requires an improved model with high-end performance accuracy in screening and predicting the SARS-CoV-2 with a different kind of related disease by analyzing the clinical, mammographic, and demographic information of the suspects and infected patients. Finally, it is evident

that AI and ML can significantly improve treatment, medication, screening & prediction, forecasting, contact tracing, and drug/vaccine development for the Covid-19 pandemic and reduce the human intervention in medical practice. However, most of the models are not deployed enough to show their real-world operation, but they are still up to the mark to tackle the pandemic.

Reference

Publication/ Screen Shots

1. Confirmed cases of Covid 19. Available from: https://www.covid19india.org/. (Accessed 26 April 2020).

2. David J Cennimo Discusses Coronavirus Disease 2019 (COVID 19). Available from: https://emedicine.medscape.com/article/2500114-overview. (Accessed 25 April 2020).

3. Wang J., Liu Y., Wei Y., Xia J., Yu T., Zhang X., Zhang L. Epidemiological and clinicalcharacteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020;395(10223):507–513. [PMC free article] [PubMed] [Google Scholar]

4. Coronavirus Disease (COVID 19) Outbreak. Available from: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-corona virus-2019-ncov. (Accessed 25 April 2020).

5. COVID 19 Article. Available from: https://www.newscientist.com/term/covid-19/. (Accessed 30 April 2020).

6. Erica Hersh Discusses How Long Is the Incubation Period for the Coronavirus?. Availablefrom: https://www.healthline.com/health/coronavirus-incubation-period#incubation-period.

(Accessed 25 April 2020).

7. Symptoms of Coronavirus. Available from:

https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html. (Accessed 24 April 2020).

8. Anulekha Ray Discusses ABOUT Coronavirus: India's biggest Concerns are COVID 19 Patients with No Symptoms. Available from: https://www.livemint.com/news/india/coronavirus-india-s-biggest-concerns-are-covid-19-patient s-with-no-symptoms-11587533159071.html. (Accessed 26 April 2020).

9. Kaggle Dataset used in this project.

10. worldometer_data.csv - Latest data from https://www.worldometers.info/

11. Acknowledgements / Data Source

12. https://github.com/CSSEGISandData/COVID-19
https://www.worldometers.info/

13. Collection methodology

14. https://github.com/imdevskp/covid_19_jhu_data_web_scrap_and_cleaning

15.

Publication/Copyright Product

The paper is not yet published but has been submitted to various journals, waiting for positive response.