Coronavirus Detection Using Machine learning

Chandana P[1], Deepashree S Rannore[1], Maheshwari B[1], Dr.Amutharaj Joyson[2], Dr.R Balakrishna[3]

[1]*Students, Department of Information Science and Engineering, Rajarajeswari College of Engineering*
[2]*Head of the Department, Department of Information Science and Engineering, Rajarajeswari college of Engineering* [3]*Prof & Dean, Rajarajeswari college of Engineering*

Abstract*: The Coronavirus Disease began on 31st Dec, 2019, in Wuhan, China, with an unknown type of pneumonia, but quickly grew into a pandemic. COVID-19 is the name of the disease, and SARS CoV2 is the name of the virus. The majority of coronaviruses afflict animals, however they can also be transferred to humans due to their zoonotic nature. SARS (Severe Acute Respiratory Syndrome) is a type of acute respiratory syndrome. Humans get infected with the Corona virus (SARSCoV) and the Serious Acute Respiratory Syndrome Corona virus (MERSCoV), which cause severe respiratory sickness and death. Asymptotic patients play a critical role in the transmission of infections. The current method for detecting the corona virus (COVID19) is RTPCR. Chest X-rays are critical in the early detection and treatment of COVID-19 illnesses. Various AI techniques have proved successful in predicting the future.*
Keywords: Coronavirus; SARSCoV; MERSCoV; Photographic Features;

I.INTRODUCTION

The Coronavirus is also known as COVID-19, is a deadly contagious viral infection caused mainly by SARS-CoV-2, or severe acute respiratory syndrome. Coronaviruses are a type of virus that can cause a variety of illnesses, including respiratory and gastrointestinal illnesses, as well as death. The initial cases of this is mostly respiratory infection which were reported in late Dec 2019 in Wuhan. Following that, the World Health Organization (WHO) declared the outbreak a worldwide pandemic on 11/3/2020, after it spread across the globe in a short period of time.

According to WHO, there have been 535,239,675 confirmed cases of COVID-19 worldwide as of December 2021, with 6,320,146 deaths. COVID-19 is disseminated widely by dust particles and fomites in close contact, as well as via touching between an infector and an infected person.

In some cases, rapid transmission has been observed, and the active virus has been detected in a small number of clinical tests. Fever, cough, headache, difficulty breathing, loss of smell and taste, and a few others are the major symptoms of Coronavirus disease. We can limit the risk of infection by wearing masks in public, keeping a safe distance from others, and frequently cleaning ourselves. The WHO continues to encourage each and every individual to take care of their own health and also protect others by following all the guidelines generated.

II.LITERATURE SURVEY

1. In 2021, Rohini M, Naveena K.R, Jothipriya G, Kameshwaran S, Jagadeeswari M contributed to the COVID-19 outbreak study, which used several Machine Learning classification results to visualise the disease's origin and perform predictions and time series monitoring to help control the disease's impact in the coming days. The disadvantage was that they could have obtained more significant results by employing a variety of other machine learning methods for estimating estimates that would have assisted clinicians, medical professionals, and government agencies in planning for real-time pandemic disease preparedness in the future.

2. Sugandh Bhatia and Jyoteesh Malhotra's paper "Nave Bayes Classifier for predicting the Novel Coronavirus" from 2021 describes the design method for predicting Coronavirus disease using the Nave Bayes Classifier in order to detect and decide whether a person has been infected with COVID-19 or not. The problem is that it was not possible to identify both coronavirus cases with symptoms and asymptotic instances, which is the only option to break the chain and regulate coronavirus propagation.

3. This advantage was given by C S Sumanth and R K Nayak in " Machine Intelligent Techniques for COVID-19 Detection: A Critical Review and Analysis " in 2021, using a chain of CT scan images mined from the COVID-19 dataset and ML algorithms such as marine predator, suspected infected patients were simplified and recovered (SIR). To get even better results when detecting COVID-19, we need to include two or more models that are compatible with the dataset and have other clinical characteristics.

4. Support vector Machine (SVM) was utilised as a Better Classification algorithm in "Severity Prediction of COVID-19 Patients using Machine Learning Classification," 2020 by H Gull, Gomathi K, M I Aldossary, and Sardar Zafar Iqbal. The necessity for the detection of COVID-19 cases with conventional symptoms was not verified, which was the study's principal issue. In 2020,

5. Akshay K Siddu, Dr. Ashok K, and Dr. Shakti K presented "Detection of COVID-19 from Medical Images and/or Symptoms of Patients Using Machine Learning . Approaches," which was primarily concerned with selecting the best DL models to detect lung segmentation and predict · COVID patients using DL techniques.

III.SYSTEM REQUIREMENTS AND SPECIFICATION

A. *SYSTEM ANALYSIS*

System analysis is nothing but a process which is responsible for grouping, interpreting and collecting the facts, identifying and solving the problems and further decomposing the system into its components. The main aim of the system analysis is to get to know about the system and its parts for the purpose of identifying its objectives and it is a technique which helps the system to work efficiently in order to accomplish their purpose. The main focus is basically on **Systems, Processes** and **Technology.**
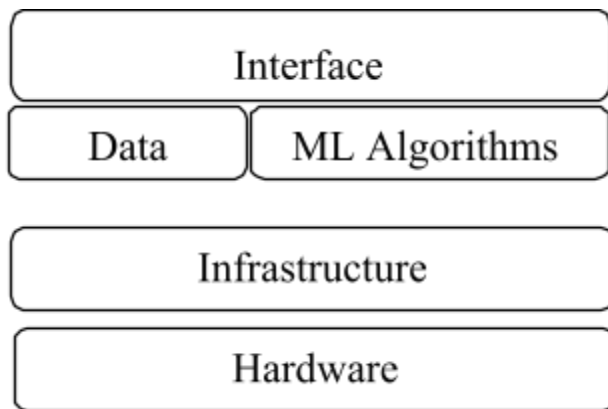
B. *FUNCTIONAL REQUIREMENTS*

In order to enable the success of a system, the requirement analysis is a very important process. The functional requirements are those where there are demands put on by the end user based on the basic facilities that the system should provide or offer. This is a type of requirement where one can see the **final product directly**.

Software Interface: An interactive UI is required for interaction between user and system designed.

Software Interface: For Modeling, Interfacing and Data extraction python packages are needed.

C. *NON-FUNCTIONAL REQUIREMENTS*

REQUIREMENTS THAT ARE NOT FUNCTIONAL: A non-functional need is one that requires the system's quality attributes to be met in accordance with the project contract. The attributes like **usability** and **effectiveness** of the system entirely is essential in a non-functional requirement. When the system fails to meet non-functional requirements, there can be possibility where the satisfaction of the user needs fails.

The non-functional requirement mainly deals with the issues like **flexibility, reliability, scalability, security** etc.

It helps in the verification of performance of the software.

- **Efficiency:** The Machine Learning algorithm prediction ability must be greater than 90% for accurate results.

- **Portability:** The system is designed to work on a computer system with low processing capacities and also independent OS.

- **Scalability:** The current system will be system dependent and hence any cloud platform can be used to scale up to any device by hosting the same system.

- **Robustness:** The data used to train the Machine Learning system should account for all data flaws.

IV.SYSTEM DESIGN

System design is one of the major stage in the process of the software development. The major and the initial goal here what the design becomes a crucial step or the first step toward a solution to a problem.

1.1 System Architecture
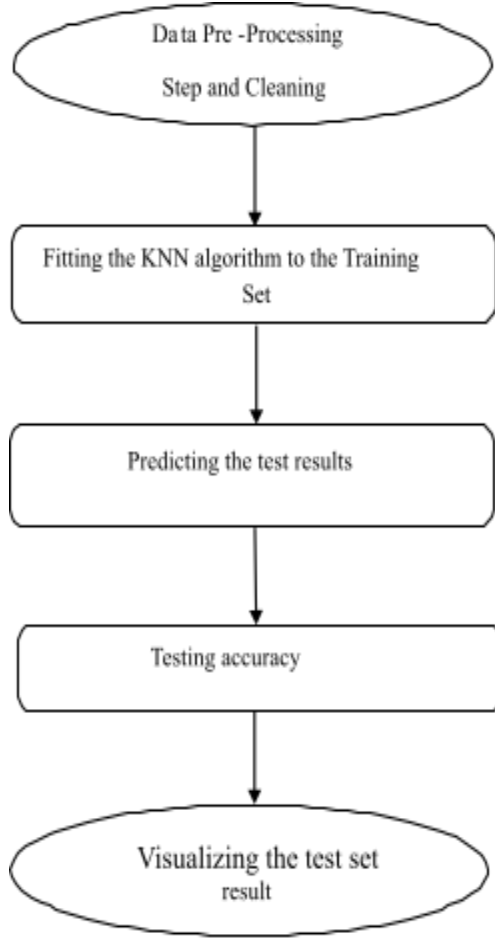
1.2 Input/Output Design:

The dataset with the required data which has to be processed has been taken as the input here where the system processes this input dataset and finally, according to the inputs, the system gives the visualized data where the data can be represented in the form of graphical images, the results predicted to a particular event or an action would happen in the future and lastly the evaluation parameters, when the sub program gets called, the actual and the formal parameters gets mapped.

V. STATISTICS OF COVID-19

There are active cases and deaths in 16 of the world's worst affected countries.

Table I: Cases of Coronavirus and Deaths by Country

Sl no	Nations	Total +ve cases	Total number of deaths
1	USA	86503057	1033571
2	India	43178080	524692
3	Brazil	31153069	667044
4	France	29621064	148644
5	Germany	26539842	139744
6	UK	22305893	178749
7	Russia	18351851	379520
8	South Korea	18163686	24258
9	Italy	17490451	166922
10	Turkey	15072747	98965
11	Spain	12403245	106797
12	Vietnam	10724554	43080
13	Argentina	9230573	128889
14	Netherland	8090237	22325
15	Japan	8929654	30735
16	**World**	**535239675**	**6320146**

Source : worldometers.info/coronavirus June 5,2022.

VI. COVID-19 PREDICTION WITH K-NEAREST NEIFGHB

o **Step-1:** Select the number K of the neighbors

o **Step-2:** Calculate the Euclidean distance of **K number of neighbors**

o **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.

o **Step-4:** Among these k neighbors, count the number of the data points in each category.

o **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.

o **Step-6:** Our model is ready.

1.1 Working Mechanism:

KNN Start

Input Testing Data

Set the value of K

Calculate the distance of Euclidian
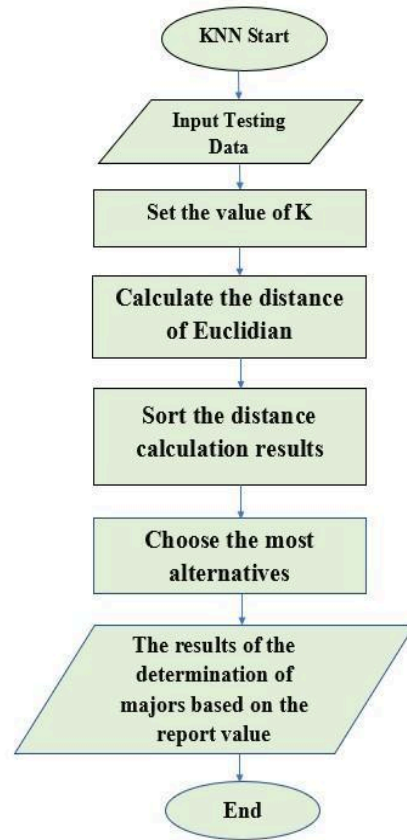
$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i\right)^2}$$

Sort the distance calculation results

Choose the most alternatives

M

The results of the determination of majors based on the report value

End

1. Determine parameter K = no of nearest neighbour usually optimum value of K is $\sqrt{N}$, where N is total no. of samples.

$\therefore$ optimum value of $K = \sqrt{1000} = 31.66$

2. Calculate distance between query instances and training sample.

P1 cough	P2 fever	P3 sore throat	P4 shortness of breath	P5 headache	P6 age >=60	P7 gender
0	1	1	0	1	1	2
0	0	0	0	1	0	0
1	0	0	0	1	0	0
0	0	1	0	0	0	0

For parameter P1 and P2

Square Distance :

$\sqrt{(0-1)^2} + \sqrt{(0+1)^2}$	$= 2$
$\sqrt{0^2} + \sqrt{0^2}$	$= 0$
$\sqrt{(0-1)^2} + \sqrt{(0+1)^2}$	$= 2$
$\sqrt{0^2} + \sqrt{0^2}$	$= 0$

3. Sort the distance and determine nearest neighbour based on Kth min distance.

P1	P2	Distance	Test result
0	0	0	Negative
0	0	0	Negative
1	0	2	Positive
0	1	2	Positive

And hence for 31 samples.

Hence, using sample majority of the category of nearest neighbours as prediction value of query instance P1 = cough = 1 and P2 = fever = 0 belongs to corona +ve category.

VIII. METHODOLOGY

A. Data Acquisition and Cleaning

Data Acquisition means collecting the data from the relevant sources before it can be stored. Data Cleaning refers to identification and correction or the errors in the datasets that may negatively or wrongly impact a predictive model.

B. Data Visualization

Data Visualization means graphically representing the data and information in a graphical or pictorial formats like charts, maps and graphs. Data Visualization enables the users to see and understand the pattern and trends in data and outliers.

C. Data Modelling

Data that has been cleaned and processed must be trained to recognise specific patterns. We teach a model to reason about and learn from datasets by providing it with an algorithm to utilise.

D. Testing

The testing method is used to assess the performance of a completely trained model. The tested data serves as a last, real-world check of previously unseen data, ensuring that the Machine Learning algorithm was properly trained.

E. Comparison and Measurement

The goal is to compare the results of the testing with those of other ML algorithms to select the best method for the data. As a result, the algorithm's performance is effective.

IX. CONCLUSION

As with any other widespread sickness, the Coronavirus Pandemic presently appears to be the most serious infected spreading epidemic. Because of the dramatic increase of COVID-19 instances during the pandemic, healthcare providers have had a difficult time coming up with acceptable and effective treatment options. Machine Leaning techniques are used as a separate strategy for viral categorization and prediction. The K-Nearest Neighbor algorithm was used to construct a supervised ML classification model for the coronavirus virus in our project. The techniques can be implemented using the KNN algorithm, which is widely used in data mining. COVID-19, also known as coronavirus. The World Health Organization has declared this a pandemic.

By implementing KNN algorithm in our project we can conclude that,

- The standard symptoms of covid-19 detection cases are verified effectively.
- Increasing the prediction accuracy to 90%+ efficiency using KNN Machine Learning technique.
- Based on the value we can easily detect multiclass prediction.
- Models like logistic regression, KNN performs better and you need less training data.
- Accurate output prediction will be useful and fruit-full for humanity and to detect the virus in the early stage and safe lot of lives.

X. REFERENCES

[1] In the year 2021, C S Sumanth and R K Nayak will publish "A Comparative Approach to Predict Corona Virus Using Machine Learning." The title of the journal is International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021).

[2] In the year 2021, S Bhatia and Jyoteesh M published "KNN Classifier for predicting the Novel Coronavirus Base Paper." Third International Conference on ICT and Virtual Mobile Networks is the title of the journal (ICICV 2021).

[3] In the year 2021, C S Sumanth and R K Nayak published "Machine Intelligent Techniques for COVID-19 Detection: A Critical Review and
    Analysis." The title of the journal is 12th International Conference on Computing, Communication, and Networking Technologies (ICCCNT).

[4] In the year 2020, H Gull, Gomathi K, May Issa Aldossary, and Sardar Zafar Iqbal published "Severity Prediction of COVID-19 Patients Using Machine Learning Classification." 6th International Conference on Computer and Communications is the title of the journal (ICCC).

[5] In the year 2020, Akshay K Siddhu, Dr. Ashok K, and Dr. Shakti K published "Detection of COVID-19 from

Medical Images and/or Symptoms of Patients Using Machine Learning Approaches." The title of the journal is 9th International Conference System Modeling and Advancement in Research Trends (SMART).

[6] Harapan, Naoya Itoh, Amanda Yufika, Wira Winardi, Synat Keam, Haypheng Te, Dewi Megawati, Zinatul Hayati, Abram L. Wagner, Mudatsir published "Journal of Infection and Public Health" in 2019.

https://www.sciencedirect.com/science/article/pii/S1876034120304329

[7] Akshay K Siddhu, Dr. Ashok K, and Dr. Shakthi K published "Detection of COVID-19 from Medical Images and/or Symptoms of Patients Using Machine Learning Approaches" in the year 2020.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnu mber=9336799

[8] Rajan Gupta, Ganrav Pandey, Poonam Chaudhary, and Saibal K. Pal, "Machine Learning Models for Government to Predict COVID-19 Outbreak," Digital Government: Research and Practice, Vol. 1, No. 4, August 2020, Article 26.