# Algorithm based Approach in Machine learning to predict the persistent spread of Death Rate of Covid 19 Variants


Dr. P. Keerthika
Department of CSE
Kongu Engineering College
keerthikame@gmail.com

Dr.P.Suresh
Department of IT
Kongu Engineering College
sureshme@gmail.com

Dr.R.Manjula Devi
Department of CSE
Kongu Engineering College
rmanjuladevi.gem@gmail.com

S. Vaishnavi
Department of CSE
Kongu Engineering College
svaish2000@gmail.com

C. Shanmathi
Department of CSE
Kongu Engineering College
shanmathi200@gmail.com

V. Surendar
Department of CSE
Kongu Engineering College
surendher21z@gmail.com



## Abstract

From closedown of December 2019, coronavirus has directly exhibited a lofty rate of transmission, coercing the World Health Organization to contend in the month of March 2020 that this unbeknownst coronavirus can be depicted as an pandemic. [1] describes this COVID-19 epidemic has guided to an operatic misplacement of deathly life over the public and presents an unbeknownst complaint to public fitness. It also affects the food systems of the person and the world of work. Once the person is infected by covid, the metabolic exertion of vulnerable cells in his or her body is enhanced, similar as the one driven by COVID-19.

In the existing research, the country's dietary habits are analyzed to predict the particular person's survival rate. By using KNN algorithm the performance metric, Accuracy is evaluated for the country's dietary habits. In this research, both clustering and classification are combined to increase the accuracy of the prediction of survival of the person. K-Means is used for the clustering of the countries and KNN for classification. The 170 countries are clustered into high and normal death rate countries based on the Countries dietary habits and another cluster into high and normal death rate based on the other disease affected rate rather than COVID - 19. Using the country’s dietary habits and other diseases affected, the survival rate of the person is predicted.

The 170 countries are clustered based on the country's dietary habits and other disease affected rate using K-Means Clustering Algorithm. After Clustering the data based on the country's dietary habits and other disease affected rate, the KNN algorithm is used to classify and identify the person’s survival rate. Using Clustering and Classification algorithms in a combined way an accuracy of 79% is achieved.


## I. INTRODUCTION

The severe acute respiratory motive coronavirus 2 contagion causes Coronavirus Illness, an transmittable disease. [2] describes the adulthood of those infected with the contagion will hold genial to moderate respiratory symptoms and will get back without the lack of medical concentration. Some, on a different angle, will pick up critically crummy and deliver medical assistance. Sedate illness is additionally probable to walk out the elderly and those with bolstering medical conditions analogous as cardiovascular indictment, diabetes, chronic respiratory indictment, or cancer. [3] says that COVID-19 may frame anyone ill and bring them to pick up truly ill or dead-end atanyage. Being fully grassed on the indictment and how it spreads is the highest plan to shirk and break down transmission. Stay at least 1 cadence downward from people, burn out a fluently decent vizard, and splash your hands or apply an alcohol- predicated annoyance periodically to pinch-hit yourself and others from infection. [4] says that when it's your wander, get vaccinated and supervise foremost counsel. Larger respiratory driblets to lower

aerosols are among the particles. [5] refer however, it's critical to borrow respiratory form, similar to coughing into a flexed elbow, If you're sick.

## II. RELATED WORKS

### A. EFFECTS OF COVID-19 ON PATIENTS WITH OTHER DISEASE

The nimbus contagion complaint 2019 epidemic has accelerated the dangers for cases with different conditions and has commanded a expressive collision on conventional health care capitals. Corresponding to primary holdings from experimental explorations, there are concrete calibers of hospitalization and ancient mortality. Patients with lung, respiratory challenges or cancer who likewise hold COVID-19 held a significantly lesser mortality rate. Older era, manly coitus, a record of smoking, the presence of multitudinous functional comorbidities, interpretation deal, and developing malice are presently verified threat deputies for mortality in people with COVID-19. [6] describes that rotundity and some attendant ails hold likewise existed recognised as probable clinical threat attorneys for mortality in people with COVID-19. During a pandemic, patients with any other disease may be a susceptible population. Because of their sickness, treatment-related adverse effects, and nutritional deficits, they are generally immunosuppressed. Furthermore, they are at a higher risk of contracting opportunistic infections and having serious complications as a result of these illnesses. Postponing or changing the treatment schedule may result in a negative consequence, with a well-documented impact on clinical outcomes.

### B. IDENTIFICATION OF COVID-19 APPLYING X-RAYS AND CT-SCANS

The use of various machine learning methods facilitate the use of X-ray images and CT-scans to diagnose covid-19 more effectively. The accuracy rate of predicting Covid-19 using X-rays and CT-scans range from 78% to more than 99%. [7] says that this method not only helps in identifying covid-19 but also helps in predicting the severity of the disease and chance of early mortality. The availability of a large database of data sets and images of X-rays and CT-scans of Covid-19 patients enables machine learning models to be trained and tested to their extent. The study provides the early identification of coronavirus-19 from X-rays and CT-scans.

### C. DIAGNOSIS BASED ON BLOOD AND URINE TESTS

COVID-19 is diagnosed using epidemiological factors, clinical characteristics, imaging results, and nucleic acid screening, among other things. These technologies took a long time to produce the diagnosis results and were prone to errors. [8] says that for a patient with COVID-19 infection, many types of clinical data were obtained and manually merged by doctors to make diagnostic judgments.Using clinical information and blood/urine test data, this study looked at detecting critically unwell COVID-19 individuals from those with minor symptoms. Age, gender, body heat, heart beat rate, vital signs, and hypertension were among the clinical data. The blood/urine tests may be performed in a technically simple and cost-effective manner. In large-scale clinical practices, an accurate severeness identification model of COVID-19 patients based on the criteria listed above may enhance the disease's prognosis.It would be feasible to follow individuals who may have contracted COVID-19 inside the hospital using daily blood samples, enabling more efficient isolation procedures. A patient would be assessed by health specialists to see if he or she is a likely case of COVID-19. Simple blood tests should be sought if the patient is identified as a suspicious case.

### D. VACCINATION

Enough vaccine boluses have nowadays been contributed to completely vaccinate 44.6 percent of the world's population, but the admeasurement has existed irregularly. [9] refers that while the best vaccinations are very successful at reducing illness and death, stopping a pandemic requires a concerted approach. Vaccinating 70 percent to 85 percent of the US population, according to infectious-disease experts, would allow for a return to normalcy.That's a frightening degree of immunization on a worldwide basis. The current global immunization rate averages 31,953,362 doses per day. It will take another 5 months to cover 75% of the population at the current rate.This progress is under jeopardy. New strains have resurrected epidemics, headed by the highly transmissible delta variant. Vaccines and viruses are now in a life-or-death battle. [10] refers that Unvaccinated persons are now more vulnerable than ever before. According to the most recent statistics, even among people who have been vaccinated, the delta variation can cause minor cases, and those who become ill can spread the disease to others. Countries have had unequal access to vaccinations and differing degrees of effectiveness in putting injections into people's arms since the global immunization effort began.

### E. NEW DELEGATES COMPANIES WITH COVID-19 TRANSMISSION

As of October 14, 2020, the COVID-19 contagion has infected over 38 million individualities encyclopedia ally, performing in over one million losses. Profitable differences, among the numerous proxies associated to COVID-19 threat, accelerated the accident of COVID-19 transmission. [11] refers that COVID-19 losses were negatively associated with per capita sanitarium bunks. Blood varieties B and AB were displayed to be defensive against COVID-19, but blood variety A was planted to be a threat agent. Downgraded COVID-19 threat was associated with the prevalence of HIV, influenza, and pneumonia. [12] refers that COVID-19 transmission is hampered additionally by lofty temperatures than by equatorial temperatures.

## III. EXISTING METHODS

In the existing research, the country's dietary habits and other disease affected person data are analyzed to predict the particular person's survival rate. By using KNN algorithm on the country’s dietary habits and the other disease affected person dataset, the performance metrics such as Accuracy, F1 score, Recall and Precision are evaluated.

K-NN is the simplest ML algorithm on Supervised Learning technique. K-NN algorithm finds the parallelism between the latest data and formerly accessible data or case and adds the latest data into the order which is additionally analogous to the data. K-NN stores already available categories and classifies the new data or case based on the similarity between data. So whenever new data comes it classifies easily based on a well suited category. The K-NN algorithm is used both for Regression and Classification. It is a non-parametric algorithm and doesn't make any assumptions on fundamental data. At the practice aspect K-NN precisely stores the data and when the latest data comes it classifies grounded on order. In figure 1 there are two categories category A and category B. A new data point arrived. The new data point belongs to category A or category B is decided by KNN algorithm based on euclidean distance.

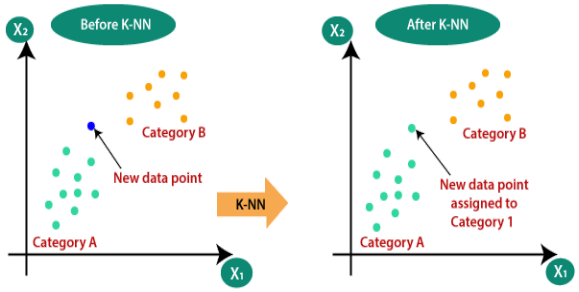


**Figure 1. KNN**

## IV. WORKFLOW

First of all, all the required dataset is collected for classification. After collecting all the necessary data, the datas in the dataset is preprocessed. By using KNN algorithm the dataset is classified. After classification, a particular person's details are given as input, to predict the particular patient's survival rate based on the country the person belongs to, the dietary habits dataset and also with the help of other disease affected dataset. From figure 2 it could be inferred the working flow of this research.

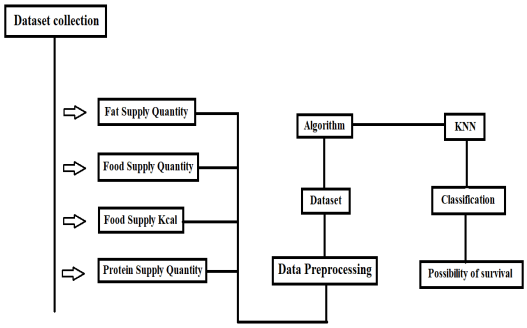


**Figure 2. Workflow**

## V. PROPOSED WORK

In this research, a hybrid of classification and clustering is used to predict the possibility of survival of the particular person. The country's dietary habits and other disease affected person data are analyzed to predict the particular person's survival rate. The country's dietary habits and the other disease affected person dataset are clustered into four types of cluster by using K-means. PCA algorithm is used to normalize the data. After clustering, the data in the dataset is given as the training data to the KNN algorithm. After training the model, a particular person's details are given as input, to predict the particular patient's survival rate based on the country the person belongs to, the dietary habits cluster and also with the help of other disease affected clusters.

### A. CLUSTERING - PCA

PCA is used to reduce the dataset's dimensionality, minimizing the information loss and increasing the interpretability. It uses matrix operations from statistics and linear algebra to calculate the

projection of original data into dimensions lesser. Dimensions of the dataset are to be reduced when there are many characteristics in the dataset, which makes distinguishing between relevant and redundant data difficult. Smaller Dataset can be easily explored and visualized, making it easier to analyze the data easily with any of the machine learning algorithms. In Figure 3 the original data is rounded in between 0 to 1.

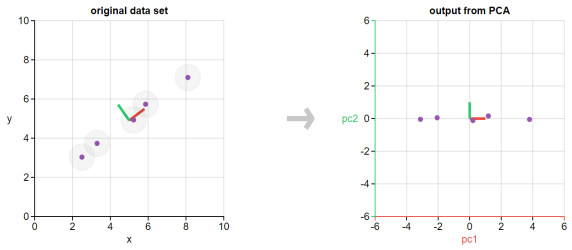


**Figure 3 Principal Component Analysis**

The foremost measure in PCA is to regularize the dataset so that each of them contributes inversely to the dissection of the data. The Reason to standardize is that if the data has larger deviation with the values then the larger value will dominate the smaller values, which will lead to biased results. Standardizing the data can prevent the problem.

$$z = \frac{(value - mean)}{standard\ deviation} \tag{i}$$

The coming shift is to pinpoint the spare data. Substantially numerous of the variables are identified in an avenue that contains spare data. To ascertain the spare facts correlation matrix is reckoned. It's a pxp symmetric matrix where p is the composition of confines.

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix} \tag{ii}$$

The coming shift is to cipher the eigenvector and eigenvalue accordingly as to adjudicate the top factors of the data. The principal components have no real meaning and are less interpretable as they are constructed as linear combinations of values. The fourth step is to find the feature vector by discarding those components of less eigenvalues. This makes the step towards dimensionality reduction. The Final step is to make use of the Feature vector formed by eigenvalues to rebuild the data represented with the eigenvalues.

$$FinalDataset = FeatureVector^{T} * StandardizedOriginalDataset^{T} \tag{iii}$$

## B. K - MEANS

The K-means is used to find groups which have not been explicitly labeled in the data. It is used to find similarities in the data.

- Intake for K-means clustering: Dataset K number of desired clusters.

❖ Affair for K-means clustering: K set of clusters.

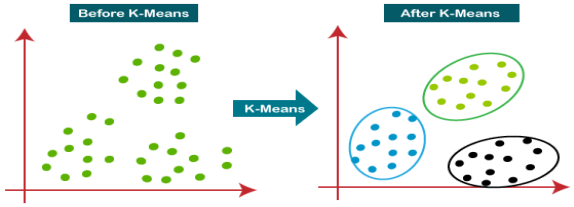


**Figure 4. K-Means**

Initialize the composition of cluster k, and gather the original centroid aimlessly. Also the squared Euclidean length will be computed from each valuation to each cluster is reckoned, and each thing is charged to the closest cluster. After that, for each cluster, the new centroid is reckoned and each value is now replaced by the separate cluster centroid. Also Euclidean length from a thing to each cluster is ciphered, and the valuation is distributed to the cluster with the smallest Euclidean length. In Figure 4 each centroid point is clustered into K clusters.

## C. CLASSIFICATION - KNN

$$Euclidean\ distance\ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \text{(iv)}$$

Step 1 : Cargo the training and Testing data set
Step 2: Handpick the values of K that is the closest data points
Step 3: For each juncture in the test data, the succeeding way are served
Compute the length between test data and each row of training data with the assistance of the Euclidean system, Manhattan system or likewise with the assistance of the Hamming system.
In this exploration, the Euclidean system is employed to compute the length between test data and each row of training data.
Next, grounded on the length valuation, the valuations are sorted in a thrusting sequence.
Currently, it will handpick the loftiest K rows from the sorted cluster,
Atlast, it will charge to the test point grounded on the most periodical order of that row.
Step 4: End

## D. WORKFLOW

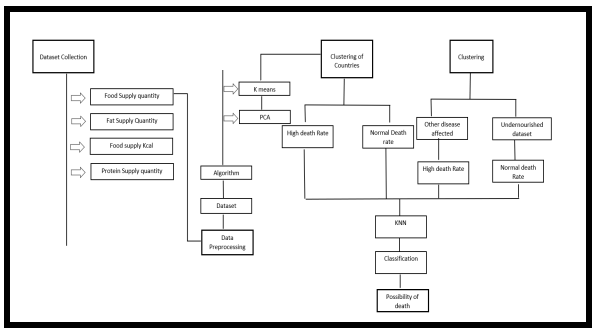


**Figure 5. Workflow**

The above figure describes the workflow of this research work. In this research, the country's dietary habits and other disease affected person data are analyzed to predict the particular person's survival rate. By using K-means clustering, the country's dietary habits and other disease affected person data are clustered into four clusters. By using KNN algorithm on the country's dietary habits and the other disease affected dataset is classified. The performance metrics such as Accuracy, F1 score, Recall and Precision are evaluated.

From the above figure it could be inferred the working flow of this research. First of all, all the required dataset is collected. After collecting all the necessary data, the datas in the dataset is preprocessed. After preprocessing, the dietary habit dataset is clustered into two clusters of high and normal death rate countries and other disease affected dataset is clustered into two clusters of high and normal death rate countries. By using KNN algorithm the dataset is classified. After classification, a particular person's details are given as input, to predict the particular patient's survival rate based on the country the person belongs to, the dietary habits dataset and also with the help of other disease affected dataset.

## VI. HYBRID APPROACH

In this research, both clustering and classification is used to increase the accuracy. For clustering, the K-means algorithm is used. PCA is used for normalizing the dataset. For classification, KNN algorithm is used. PCA is the important method for reducing the dimensionality of the dataset. The K-means is used to find groups which have not been explicitly labeled in the data. It is used to find similarities in the country's dietary dataset and other disease dataset. By using both k-means and KNN, the survival rate of the particular person was predicted. Compared with existing work, the hybrid approach, the performance metrics, accuracy was increased.

## VII. DATASET

It combines data of dissimilar classes of food and COVID-19 cases and deaths each around the people. Each dataset contains data about 170 county's food practices . The dataset includes % of fat consumed, % of food supply (in kilogram) consumed, % of energy (in kilocalories) consumed, % of protein consumed. The Fat consumed dataset, food supply in kg dataset, food supply in kcal dataset, and protein consumed dataset contains animal products, aquatic products, fish, seafood, fruits, vegetables, meat, oil crops, pulses, cereals, species, and also COVID-19 active cases, death cases, recovered cases and population. The dataset also contains the details of the person affected with any other disease other than COVID - 19. It includes diseases like Diabetes, CKD, COPD, Obesity and Undernourishment. Table 3.1 describes the cluster. Countries are clustered into high death rate and normal death rate countries based on other disease affected person dataset.

## VIII. PERFORMANCE METRICS

Delicacy (Accuracy) is a metric for assessing bracket miniatures. Informally, delicacy is the bit of prognostications the miniature was accurate. Formally, delicacy is the number of accurate prognostications per total number of prognostications. The Delicacy attained is

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad \text{(i)}$$

In equation (iv), TP is the number of true positives, FN the number of false negatives, TN is the number of true negatives, FP is the number of false positives.

## IX. PERFORMANCE EVALUATION:

Figure 7 describes the accuracy for KNN Classification and the Hybrid Approach. In the y-axis the percentage is taken and in the x-axis accuracy for KNN Classification and Hybrid Approach was taken. Figure 8 describes the precision for KNN Classification and Hybrid Approach. In the y-axis the percentage is taken and in the x-axis precision for KNN Classification and Hybrid Approach was taken. From the below graphs it is clear that the proposed system performs more efficiently than the existing system. The accuracy of the proposed system is 79 percent.

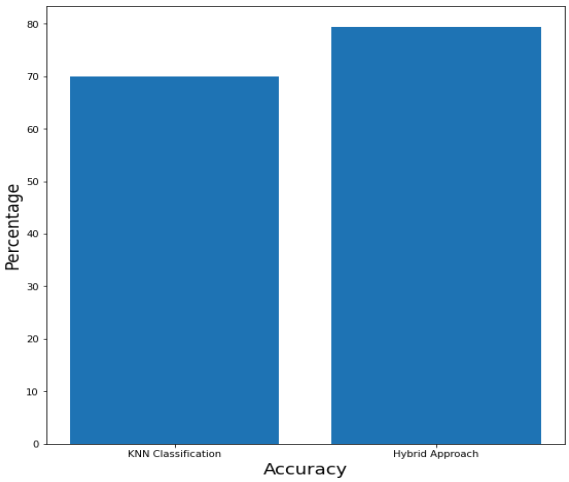


**Figure 7. Accuracy**

## X. CONCLUSION AND FUTURE WORK

The goal of this project is to predict the person's possibility of survival based on the country they belong to and if the person is affected by any other disease. Narinder Singh Punn, (2020) refers that Covid-19 is being spread rapidly and many technologies have been identified for prediction of survival rate. KNN is the most commonly used classification algorithm. Though many inventions have been made, using KNN algorithms is the best and fastest way to predict the survival rate of people. It stores all the data that is fed through clustering of the data based on country dietary habits and other diseases and classifies the new data, based on similarity in the test data and training data. The purpose of this project is to predict the person’s possibility of survival.

The future work of this project is to develop a web based application which will be useful for the hospital management system and make it free of cost. And have planned to develop two additional modules, assessment using RNN and IoT. The analysis using RNN module will act as a voice enabled chat bot and ask some questions to the users. It is intelligent to change the questions from the user’s mental state and their previous answers. The final module is analysis using IoT which makes use of pulse rate monitors and EEG sensors to detect depression at a very accurate level. This will be very effective and useful for the doctors and also the public.

## REFERENCE

1. Geza Halasz, Michela Sperti, Matteo Villani, Umberto Michelucci, Piergiuseppe Agostoni, Andrea Biagi, Luca Rossi, Andrea Botti, Chiara Mari, Marco Maccarini, Filippo Pura, Loris Roveda, Alessia Nardecchia, Emanuele Mottola, Massimo Nolli, Elisabetta Salvioni, Massimo Mapelli, Marco Agostino Deriu, Dario Piga, Massimo Piepoli (2021) A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score. Journal of Medical Internet Research, Vol 23, issue No: 5, doi: 10.219
2. Ameer Sardar Kwekha-Rashid, Heamn N. Abduljabbar, Bilal Alhayani (2021) Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Applied Nanoscience, doi : 10.1007/s13204-021-01868-7
3. Francesca De Felice, Antonella Polimeni (2020) Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis. in vivo 34: 1613-1617 (2020), doi : 10.21873
4. Mustafa Abdul Salam ,Sanaa Taha ,Mohamed Ramadan (2021) COVID-19 detection using federated machine learning. Plos one, doi : 10.1371/journal.pone.0252573
5. Furqan Rustam; Aijaz Ahmad Reshi; Arif Mehmood; Saleem Ullah; Byung-Won On; Waqar Aslam; Gyu Sang Choi (2020) COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Access, Vol 8, Page no : 101489 - 101499, doi: 10.1109/ACCESS.2020.2997311
6. Gergo Pinter, Imre Felde, Amir Mosavi, Pedram Ghamisi, Richard Gloaguen (2020) COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. MDPI, Issue no : 6, Page no : 890, doi : 10.3390/math8060890
7. Sakifa Aktar, Md. Martuza Ahamad, Md. Rashed-Al-Mahfuz, AKM Azad, Shahadat Uddin, A H M Kamal, Salem A. Alyami, Ping-I Lin, Sheikh Mohammed Shariful Islam, Julian M.W. Quinn, Valsamma Eapen, Mohammad Ali Moni (2020) Predicting Patient COVID-19 Disease Severity by means of Statistical and Machine Learning Analysis of Blood Cell Transcriptome Data. JMIR Med Inform 2021, Vol 9, Issue No 4, Page No e25884, doi : 10.2196/25884
8. Soares, F. (2020) A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams. medRxiv, doi : 10.1101/2020.04.10.20061036
9. Nathan A. Brooks, Ankur Puri, Sanya Garg, Swapnika Nag, Jacomo Corbo, Anas El Turabi, Noshir Kaka, Rodney W. Zemmel, Paul K. Hegarty, Ashish M. Kamat (2021) The association of Coronavirus Disease-19 mortality and prior bacille Calmette-Guerin vaccination: a robust ecological analysis using unsupervised machine learning. Sci Rep 11, 774 (2021). doi : 10.1038/s41598-020-80787-z
10. Nasiba M. Abdulkareem, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, Dathar A. Hasan (2021) COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms. Qubahan Academic Journal, Vol 1, Issue No 2, Page No 100–105, doi : 10.48161/qaj.v1n2a53
11. Richard F. Sear, Nicolás Velásquez, Rhys Leahy, Nicholas Johnson Restrepo, Sara El Oud, Nicholas Gabriel, Yonatan Lupu, Neil F. Johnson (2020) Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning. IEEE, Vol 8, Page No 91886 - 91893, doi : 10.1109/ACCESS.2020.2993967
12. Piu Samui, Jayanta Mondal, Subhas Khajanchi (2020) A mathematical model for COVID-19 transmission dynamics with a case study of India. Chaos, Solitons & Fractals, Vol 140, doi : 10.1016/j.chaos.2020.110173