RESEARCH ARTICLE

SAMRIDDHI Volume 15, Issue 3, 2023 Print ISSN: 2229-7111 Online ISSN: 2454-5767

An Influential Remedy in Controlling of COVID-19 based on Predictive Models of Machine Learning: A Novel Spec

Sunil Kumar*, Poonam Phogat

Department of Computer Science and Engineering, Global Institute of Technology And Management Affiliated to Maharshi Dayanand University Rohtak Haryana, India

Abstract

Coronavirus (COVID-19) is a highly potent virus that infects healthy living cells. COVID-19 the virus, makes copies in the organs of the body that replicate itself, leading to the death of some healthy people's cells and as a result, weakening the immune system. In the soft tissue, it greatly affects the respiratory tract and leads to pneumonia, organ failure, and death, reaching its final stages. This paper focuses on the early detection of COVID-19 patients based on the good symptoms of the disease. In this paper, the COVID-19 Remedies (CURE) program is proposed on a machine-based basis, reading speculation models for the treatment of COVID-19 patients. With test results, performance analysis of The CURE system is being tested in a Python environment that is tested using the Kaggle database of Johns Hopkins University.

Keywords: COVID-19, Prediction Model, Machine Learning.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2023); DOI: 10.18090/samriddhi.v15i03.03

Introduction

The old virus COVID-19 is a very difficult respirator coronavirus-2 (COVID-2) syndrome was the first of its kind was found in late December 2019 during the investigation outbreaks in Wuhan, China. The cases were rapidly growing worldwide, says WHO epidemic on March 11, 2020. COVID-19 transmission has become uncontrollable because the number of cases has reached the threshold limit.[1] The virus enters healthy living cells body and make copies of parts of the body in the last recurrence led to the death of some cells are healthy and as a result, weakens the immune system In the soft phase, it greatly affects breathing, pneumonia, organ failure, and death to reach the final stage.[2] The disease is severe in old age people with weakened immune systems and already having other chronic diseases such as diabetes, which are high blood pressure, and cardiovascular disease.[3] Figure 1 shows global statistics up to July 30, 2020, in all confirmed cases, active cases, in total deaths, and cases completely cured of the COVID-19 virus. Picture 1 (a) reveals the total number of coronavirus cases in different countries, indicating that high-profile cases followed the virus's rapid spread in the USA in India. Total number of verified cases throughout 2,18,69,976 landowners, of which 26,47,663 are Indians. Figure 1 (b) shows the performance statistics cases, of which 65,04,303 active cases have been committed worldwide. Figure 1 (c) introduces the number of death cases, and finally, Figure 1 (d) shows the number of cases treated.[4] This is a the virus that spreads communication, spreads respiratory drops are present in the air. These aerosols arrive in an open area where an infected person is sneezing and coughing, getting into other people's mouths and nose and reach to the lungs. There is no Accurate treatment to treat COVID-19. Some steps are taken to eradicate the virus using various drugs such as Hydroxychloroquine, an anti-malarial drug. Currently, it is used to treat patients with coronavirus, which helps prevent infection by increasing endosomal pH that gives enough energy to the immune system the immune system.[5] Further prevention is needed in the treatment this epidemic. Since the start of COVID-19, governments in almost all countries have taken over strong actions such as total closure, social isolation, sanitizer, and masks to reduce the whole cause elements.[6] By examining various subjects, Machine Reading seems to be the best predictive model for predicting an increase in COVID-19 cases. Descending and how to differentiate ML function according on the availability of diagnostic data for this problem

Corresponding Author: Sunil Kumar, Department of Computer Science and Engineering, Global Institute of Technology And Management Affiliated to Maharshi Dayanand University Rohtak Haryana, India, e-mail: sunilkumar141290@gmail.com

How to cite this article: Kumar, S., Phogat, P. (2023). An Influential Remedy in Controlling of COVID-19 based on Predictive Models of Machine Learning: A Novel Spec. SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 15(3), 291-300.

Source of support: Nil

Conflict of interest: None

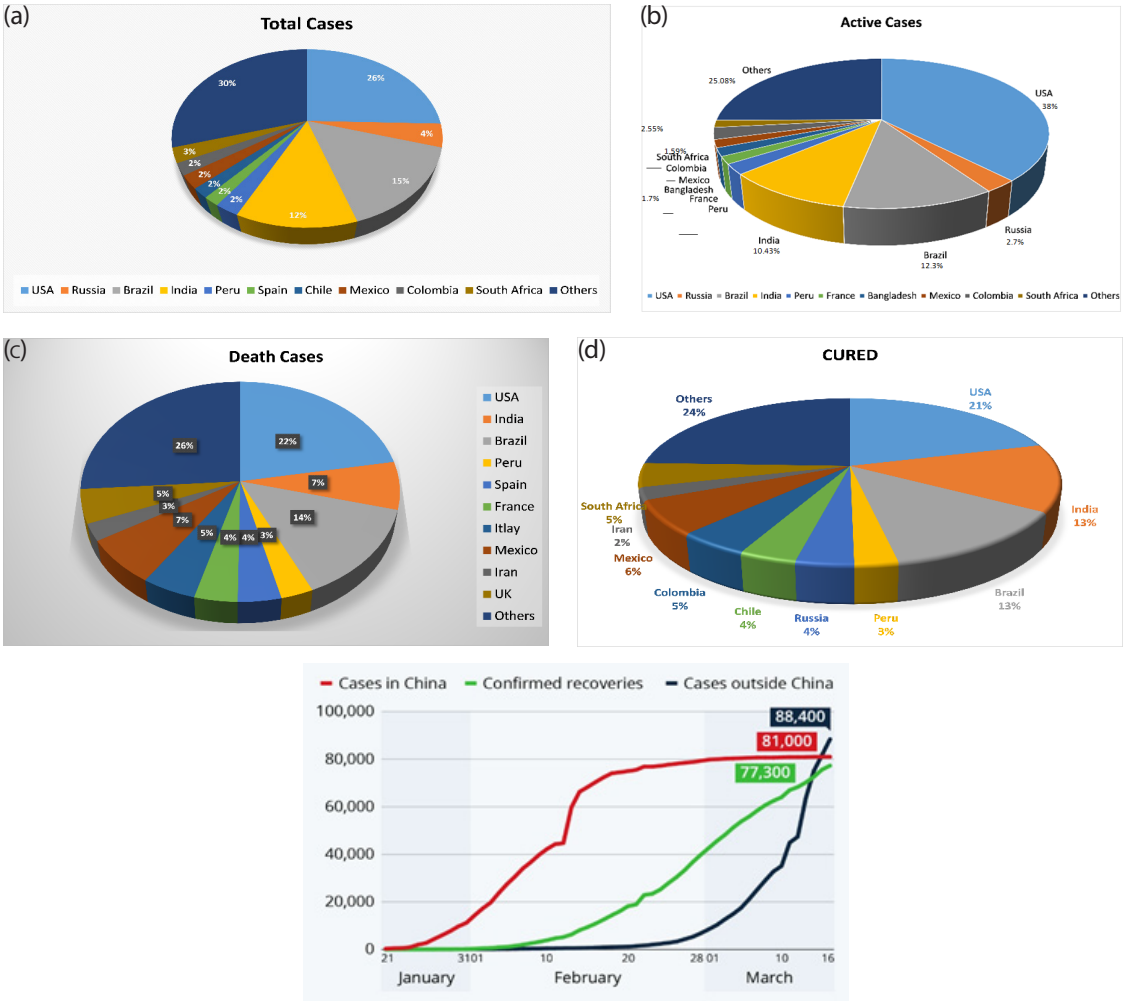


Figure 1: Data statistics of total, active, death, and cured cases on COVID-19

Table 1:

	Active Cases	*Total Cases*	*Total Deaths*	*Total Cured*
India	6,76,900	26,47,663	50,921	19,19,842
India's Share(%)	10.4%	12.1%	6.6%	13.3%
World	65,04,303	2,18,69,976	7,73,741	

Contributions

Paper contributions are summarized below.

- Identify patients' symptoms of COVID-19based on the identification of diseases.
- To rehabilitate COVID-19 patients, the CURE system is a proposed machine-based study forecasting model best prediction treatment of COVID-19 disease.
- To imitate, the proposed system is evaluated using the Kaggle database.
- Finally, performance appraisals are compared in five stages and predict the most efficient result is using the Python platform

Paper Organization

The whole paper is as follows: Phase II discusses a document review of existing programs. Phase III introduces a system model followed by a proposalCURE Plan in Phase IV. Section V contains for the performance test of the forecast, and finally, Section VI concludes the paper.

Literature Review

Researchers introduced other Machine methods Learning to distinguish. Very simple separationLinear Regression method used for reduction total square difference between real and predicted data. Problems with this model are its inefficiency with inconsistent data and sensitivity in deviation.[7] With the Logistic Regression model,it is shown that the appearance of the end is Logisticjob-based. The beauty of this model is that has no problems. But it failed to take the line. The Naive Bayes Model proposes to close training data to calculate inevitable and effective parameters dealing with real world data. Another modelK-Nearest Neighbor, shows it works well with modest and consistent data for multiple classes. [8],[9] Pinter *et al.* [10] Proposed Mechanical Learning Methodsa competitive parallel

perceptron-imperialist algorithm(MLP-ICA) and flexible network-based ambiguous interferenceCOVID-19 forecasting system (ANFIS)confirmed cases and deaths. This model is used to maintain the accuracy of the next 9 days it provides convincing results.[11] Government and society they should inform researchers and help reduce itdata by maintaining social distancing and tracking other safety measures.[12] Hamzeh *et al.* [13] is active Susceptible-Exposed-Infectious-Recovered (SEIR) Model which predicts good performance in central data. Outbreaks appear to be exacerbated during pregnancy and in childbirth in data prediction. Jia *et al.* [14] describe four phases of COVID-19 cases.In the first stage, the human travel history has COVID-19 markers leading to closure. The virus reaches the second stage when an infected person comes in contact with others in humans. To prevent using growing social distance data. Next,the third stage where there is no travel history or contact with an infected person. So chances are the spread of the virus through respiratory droplets into up. Therefore, the use of mask and sanitizers is required.The next and final stage is the uncontrollable phase, where cases have reached a limit. Tuli *et al.*[15] improved COVID-19 prediction using the model of Machine learning. In this model the data is processed used to help the government and the community. After covering data with ML and AI, researchers can predict the scale the time and regions where it is possible to distribute of this high disease. This is predicted using different ML models, COVID-19 cases can be is controlled or removed from all countries a country facing this crisis.

System Model

Figure 2 shows the workflow of the proposed KALAPHICOVID patient treatment program. Initially, the input data taken from Johns Hopkins University Database. Then the signs of good cases are analyzed and divided into 3 parts: Difficult, Medium, and intermediate symbols. Very painful patient symptoms involving throttling have to deal with severe time. Moderate symptoms include shortness of breath, fever, cough. Mild symptoms include fever, cough, and a headache. The proposed COVID-19 outbreak program analysis is trained and evaluated in real-time data using I COVID-19 patient symptoms of COVID-19. The problem is exacerbated by data inequality. Kumedical data problem class inequality is common what happens with the rule of many cases to others classes over others. Several definitions are appropriate for both algorithmic and data rate to manage unequal data. In this paper, the performance of 5 stages and a decrease in relative to an unequal database obtained while studying the predictions of COVID-19. According to the findings of this retreat and phases, the impact of SMOTE (Synthetic MinorityOversampling Technique) - an effective method and an unequal set of data, carefully tested. With the convenience of the algorithms used in this method,k samples find out which ones are nearby small samples in small and standard classes. The Euclidean distance method is used to find this distance.With the number of cases in small and large quantities classes, an unequal set of data is taken. Based on-separately independent, the original database is separated into two sets - training set (80%) and test sets (20%) used random samples with stratified. By using SMOTEstrategies, training set is more than samples to obtain the distribution of the best class in the database and 8 training sets find between them that 1 is the original set out of 7 on a sample set with different levels.

Proposed Treatment Plan

The proposed CURE system uses many different methods and tools used for forecasting. In combination of different models-SVM (Vector Support Machine), LR(Line Down), k-NN (k- Close Neighbors), EditingNaïve Bayes tool and R, learning machine the model is proposed to predict the COVID-19 infection rate. The collected Data set is cleared before further processing and is considered a first step in obtaining information on the website. For typographical problems, this data purification process is performed using Machine learning strategies. Process using methods finding inaccurate and incorrect data, correcting errors and checking the basics of information is called data cleaning; assembling and disassembly of data is involved. Data cleaning is performed on most aggregated data what the appearance of duplicate records does. Four features that are guaranteed to include quality, accuracy, integrity and consistency. The main symptoms of this disease include loss of taste and odor, headache, fever, dizziness, fatigue and shortness of breath. As critical, the symptoms are categorized into three categories: soft, medium, and solid. Mild symptoms include fever, cough, headache. The frequency of sensitivity is low at this stage. Then it reaches the middle stage, where the exhaustion main symptoms and high fever and cough. In the critical stage, the patient reaches a critical state and it becomes very serious. Respiratory problems a major problem patients have to deal with. The virus strongly affects lung damage to alveoli, transporting oxygen to all parts of the body through blood vessels and RBCs, respectively. The virus is a harmful wall of the alveolus and results in its stiffness due to which air transfer to RBCs reduces, eventually leading to hypoxia. Due to lack of food and oxygen, the chances of organ failure are always high. Collected data is first trained and tested using different models- SVM (Vector Support Machine), LR (Line Delay), k-NN (k- Nearby Neighbors), divorce and Naïve Bays. The explanation for these predictions is as follows:

Linear Regression (LR)

LR is a widely used mathematical method of prediction analysis in Machine Learning. Based on surveillance reading, a drop-down is a machine learning algorithm which performs the retrieval function. LR prediction model use the given data points to find the right oneline of measure to train the database.

Simple line number

says $y = mx + c$,

where y is a depend`ent variable, x independent variables, and m, c do not change their values

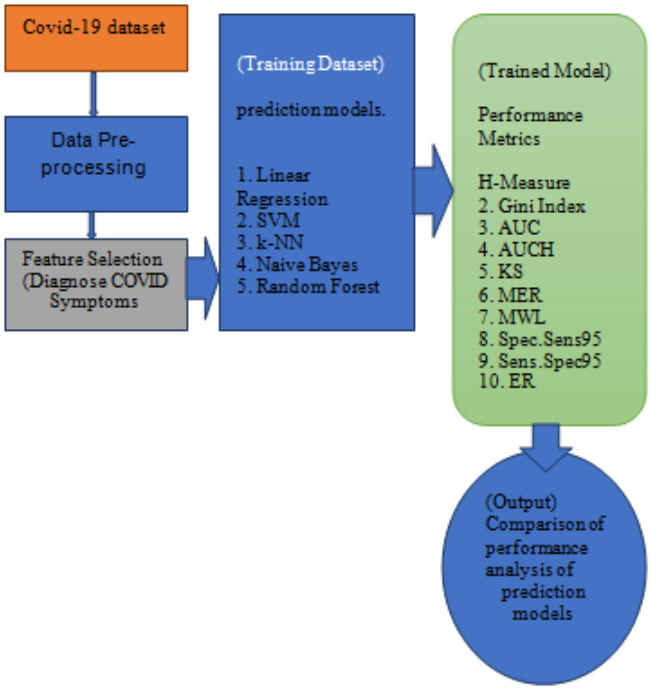


Figure 2: Workflow of the proposed CURE Scheme for the treatment of COVID patient.

they are calculated using calculation theories. Figure 3 (a) shows an example of a predictive LR predictive model features such as input and continuous output prediction as a result of finding a curve that is the line of a particular problem. The output of the LR model is calculated using number.

$y = \mu 0 + \mu 1x1 + \epsilon$, (1)

where $\mu 0$ represents y crossing, $\mu 1$ represents slope,

$x1$ input value, ϵ represents the error name, and y states model output value. Early on training, β starts at random but we prepare μduring the training specified in each aspect that loss (deviation from what you want and predictedoutput) reduced. Loss metrics are calculated by using mean squared error (MSE). These are the benefits of using LReasy implementation: easy, fast, standard training to avoid overload, and easily updated with new data using gradient decrease. The disadvantage of the LR model is that it does not work well in offline relationships, does not changeto capture complex patterns, and can be time-consuming eating. However, different output output i.e., 0 or 1, the retrospective model (binary categories) is used. Figure 3 (b) shows an example of Logistic retreat, calculates the total amount of input variables similar to the LR model but uses the output indirectly sigmoidal function to produce output.

$y = 1$

$1 + e - x$, (2)

when x is the input value, y the output value ofmodel, and e is powerful. The predictable LR model can be used in Python.

Vector Machine support method (SVM)

SVM is a ML-monitored algorithm used for both classification nand decline. Example SVM classifier shown in Figure 3 (c) representing of different categories in the decision plane or hyperplane inn-dimensional space. In this figure, the supporting vectors data points near the hyperplane. These data points are divided into classes using division line ($H1$, $H2$, $H3$). Here, the margin is defined as a gap a perpendicular distance from the line to the support vectors. SVM aims to separate databases in classes to calculate the central hyperplane. Initially, SVM detects repeatedly isolated hyperplane classes based on that SVM prefer hyperplane that divides classes best. SVM can work well on non-linear partitions while running line separation. With larger spaces and situations with a magnitude greater than a number of samples, very effective. SVM conversion input vector n-dimensional space known as element space (f) using non-line function and then linethe line deflection function is performed in space. Icon applied to Python using SVM kernels. TypesSVM kernels are linear kernels, polynomial kernels, and radial bias function (RBF) kernel. Linear Kernel: It is a dot product between two views and line kernel function is defined as using equation.

$f\ (v, vi) = sum\ (v*vi)$, (3)

where v, vi there are two vectors.

Polynomial Kernel: Curved or indirect bias

Input space defined using a calculation.

$f\ (v, vi) = 1 + sum\ (v*vi)\ d$, (4)

where d is a polynomial level made by hand set to learning algorithm.

Radial Bias Function (RBF) Kernel:

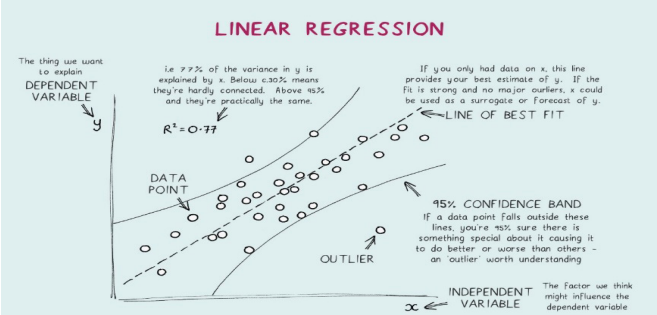
Converts inputspace in a multi-dimensional space defined asusing equation.

$f\ (v, vi) = exp\ (-\gamma*sum\ (v*vi)\ 2)$, (5)

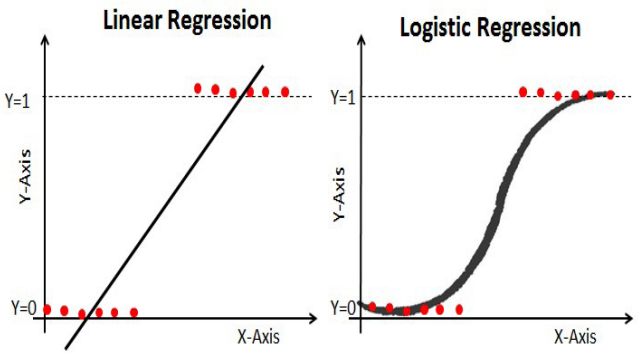
when γ is between 0 and 1, manually set and its default value is 0.1.Steps to follow in using SVM separator according to the paragraphs of text they are as follows: (i) enter svmpackages. (ii) upload the input database. (iii) select features from the database. (iv) a section of the original SVM boundaries data. (v) generating values of the stop parameter. (vi) The SVM classifier object was created using a kernel (linear, polynomial, RBF). (vii) record keepingseparation. The advantages of using SVM separators are high accuracy with multi-dimensional space, very stores a little memory and use a set of training points. The disadvantage of SVM categories is that they work SVM does not measure on large databases due to height training time, and it does not do well in the break classes. Therefore, the pruning tree is generally preferred over SVM in large databases.

k-NN (k-Nearby Neighbors)

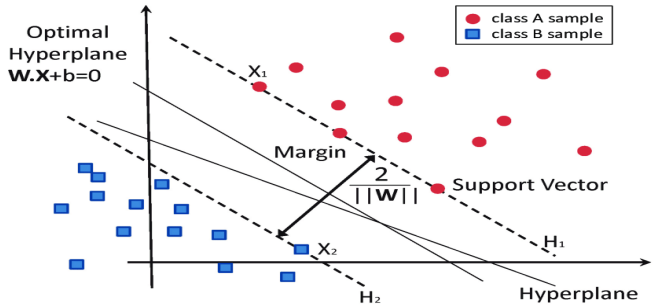
k-closest (k-NN) ML algorithm monitored a process often used for separation problems.It can be used for both split and single retreat. k-NN method divides documents based on similarity estimates measuring features as distance and proximity, parallel two data points are calculated and classified based on proximity neighbors for each data point.



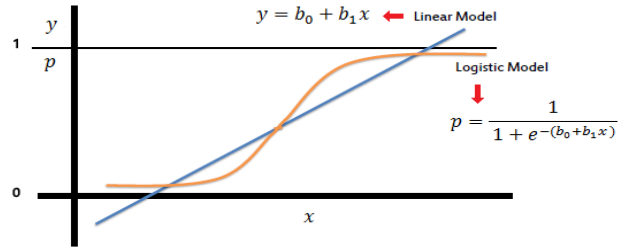
(a) Linear regression model



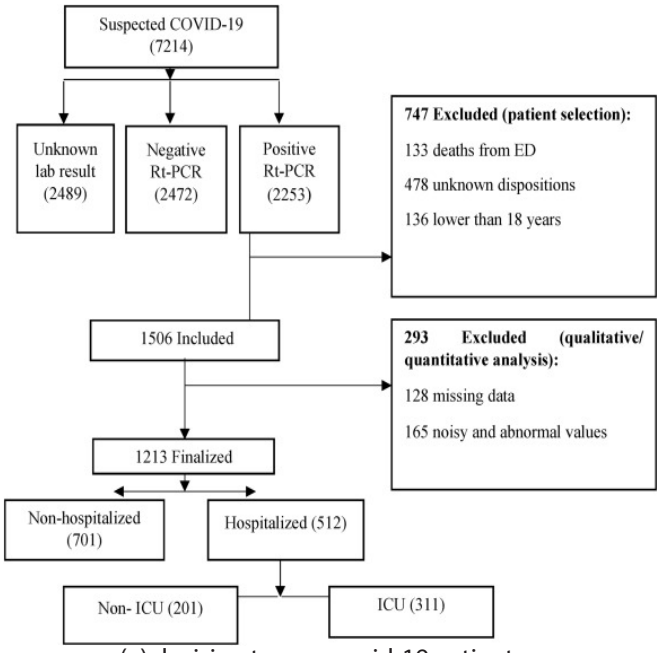
(b) Logistic regression model



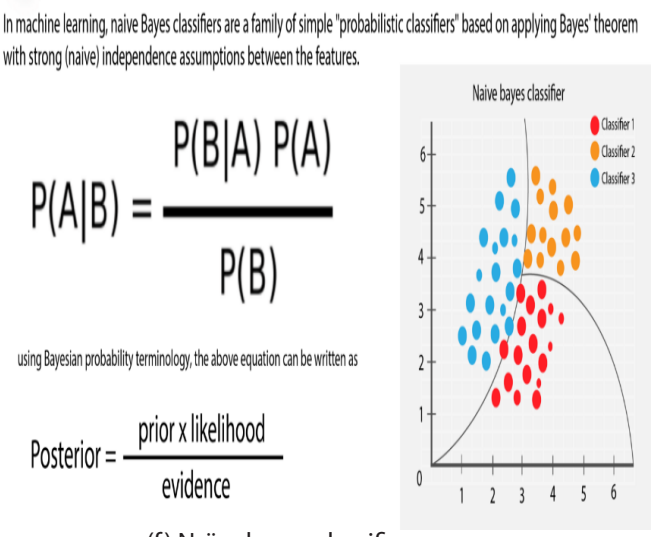
(c) SVM model



(d) K-NN model



(e) decision tree on covid-19 patient.



(f) Naïve bayes classifier

Figure 3: consists of figure (a),(b),(c),(d),(e) and (f).

Figure 3 (d) shows I an example of a k-NN model that captures intimacy two data points (same data points). K-NN is active on the principle of feature similarity to predict prices for new data points. So, a new data pointassign a value based on proximity as it is data points in the training set. Steps involved in the k-NN algorithm is as follows: (i) Upload training once database testing. (ii) Select the value of k (number) i.e. I data points are near. (iii) For each point in the test data, calculate the distance between the test data and each line of training data with the help of Euclidean or Haming distance and arrange distance values in an ascending manner.(iv) Select the k-lines at the top of the list. Next, assign a class to the test area based on multiple timessection of these lines. (v) the end result. The k-NN algorithm can be used in Python by approx. by applying the following method: (i) the required import python, (ii) download Kaggle COVID-19database, (iii) assign column names to the database, (iv) readDatabase in pandas data framework, (v) performs preliminary data processing,(vi) separate data from train and test databases (60%training data and 40% of test data), (vii) generate data measure, (viii)

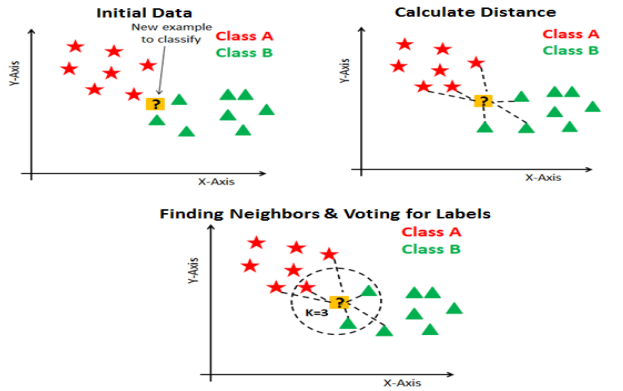


Figure 4:

train model using neighbors close to Ksklearn phase, (ix) find prediction, (x) output results- matrix of confusion, split report, and accuracy. The advantages of k-NN algorithms are simple, useful for indirect data, high accuracy. Limitations of the k-NN algorithm that the algorithm is as expensive as it saves all training data. In addition, it requires additional memory retention, and prediction slows down when there is a large database.

Naïve Bayes

Naïve Bayes is a Bayes-based approach theorem that works with the goal of solid speculation conditional independence that the presence ofa feature in the classroom is independent of the presence of any another feature in the same class. Let us consider an example for smart 4K TV, smart TV is considered asa smart category when combining features such as internet connection, high resolution, Bluetooth, USB ports, HDMI connection, support multiple applications. However, these depend on one another but each aspec tto give independent to intellectual opportunities 4K TV is a smart TV. The Naïve Bayes are very dangerous. An algorithm that can be trained on a small database.Figure 3 (e) shows an example of the Naïve Bayes model separate data points based on background opportunitiesclass into three different classes i.e., phase 1 (red data points), split 2 (orange data points), and separator 3 (blue data points). The saying of the Naïve BayesThe algorithm based on the Bayes theorem is defined as follows.

$$(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (6)$$

where $P\ (A \mid B)$ indicates the posterior possibilities section, $P\ (B \mid A)$ indicates the probability of the presence of a prediction given category, while P (A) refers to pre-class opportunities,and P (B) refers to the peripheral or previous possibilities of the predictor. Building a prediction modelusing the Naïve Bayes classifier, the model is classified into three types: (i) Gaussian Naïve Bayes (GNB), (ii)Bernoulli Naïve Bayes (BNB), and (iii) Multinomial NaïveBayes (MNB). Python Library, Scikit Learning is the largest a useful library that helps us build the Naïve Bayes modele Python. We have the following three types of NaïveBayes model under Scikit read Python library.GNB Classifier: Based on that assumption data from each label is taken from simple Gaussian distribution. MNB Classifier: Here, features are availableit is thought to be derived from a simple Multinomial distribution which is very suitable for the features that represents different numbers. BNB Category: Consider BNB features that will be binary (0s and 1s). For example, text classification by model 'word bag'. Steps involved in applying the GNB category toPython is as follows: (i) import GNB packages belowScikit read Python library. (ii) to obtain portions of pointsusing the *make_blobs* () Scikit function inGaussiandistribution. (iii) in the GNB model, we need to import

GaussianNB and do its thing. (iv) make predictionsafter receiving new data. (v) edit new datato find its limits. (vi) use line of code to calculateopportunities behind labels. (vii) output list. Ithe benefits of using the Naïve Bayes classifier are quick and easyimplementation, small training data, meets faster than thatDiscrimination models are like retroversion, and they are appropriateyour both continuous and separate data. Ithe boundaries of the Naïve Bayes section are usually zeroin case the variable is assigned to a category but notseen in a set of training data, followed by the Naïve Bayesthe editor sets zero opportunities and does not provide predictions,add autonomy as in real life app as it isdifficult to have a set of completely independent featuresto each other. Applications for Naïve BayesCategories are real-time predictions, multi-phase predictions,text separation.

Decision Tree Induction Classifier

is a simple, easy-to-understand separator that is not a parameterbased on flexible tree algorithm.It can do both split and back withhelp algorithms used to make this model fromoriginal database, random selection for trainingdata complete. Steps to get involved inThe functionality of the decision tree algorithm is as follows. (i)selection of random samples to a specific database. (ii)create a decision tree for all samples and calculationspredictive effect on all decision trees. (iii) votingperformed on all predicted results. (iv) very selectivepredictive outcome voted as predictive outcomealgorithm.Decision tree is used in Python throughthe following methods. (i) import is requiredPython packages, (ii) download Kaggle database, (iii)assign columns to the database, (iv) read the data set from it pandas dataframe, (v) performs preliminary data processing by using script lines, (vi) divide data into train and test splt (say, divide the database into 70% training data again 30% of test data), (vii) train model tree model with the help of RandomForest Classifier class of sklearn, (viii) generate guessing using text, and (ix) retention the output is a confusion matrix and a split report. Figure 3 (f) shows an example of a law based on three key features of the patient's database of COVID-19 namely, lactic dehydrogenase (LDH), high C-reactive sensitivity roteins (hs-CRP), and lymphocytes. Decision tree obtained by random classification of a total of 600 patients the root of the forest which is the number of patients in order to training and validating data sets, while the leaf area returns the result as the number of

cures and death patients. The main advantages of using the decision tree model are is worth a large range of data sets, overcoming I the problem of overcrowding by combining different results cutting trees, flexible and with very high accuracy, data measurement is not required. Decision limits The drug algorithm is highly complex, complex and complex time consuming compared to other speculative models, and requires more calculation resources.

Performance of Predictive Models

TestingThe effectiveness of speculation models can be tested as a variety of metrics labeled as follows:

(1) H-measure, (2) Gini-Indicator, (3) Location Under the Curve(AUC), (4) The convex Hull area of the ROC Curve(AUCH), (5) Kolmogorov-Smirnoff (KS) Statistics, (6) MinimumError Rate (MER), (7) Minimum Cost RatedError Rate (MWL), (8) Specification when Sensitivity is detectedfixed at 95% (Spec. Sens95), (9) Sensitivity when Specificationheld fixed 95% (Sens.Spec95), and (10) ErrorRate (ER).H-ratio: H-ratio is an

important measureclassical performance that measures accuracyof the model. The main figures of interest arethe so-called mis-classification statistics, that is, the number ofFalse False (FN) and False Good (FP). Bangufour conditions in predictable modeling. (i) Good indeed(TP): In the case of true traction (TP), the real thing is goodand it is predicted as good. (ii) False positive (FP):In the case of false positives (FP), the real one is a single objectionare predicted as good things. (iii) False False (FN): Incase of false negatives (FN), in fact there are good and therepredicted as negative. (iv) Real negatives (TN): If possibleof true injustice, the real evil is evil and foretoldas positives. An example of false confession of eventswhen the disease is diagnosed by mistake, and for examplefalse opposition events in the presence of ahealth is not available. Accuracy (AC): Accuracy in data sets provided withdata points (TP + TN) is a measure of complete predictionby dividing by the number of data points. Priceof AC is between 0 and 1.

$$Ac = \frac{(TP + T)}{(TP + TN + FP + FN)} * 100 \qquad (7)$$

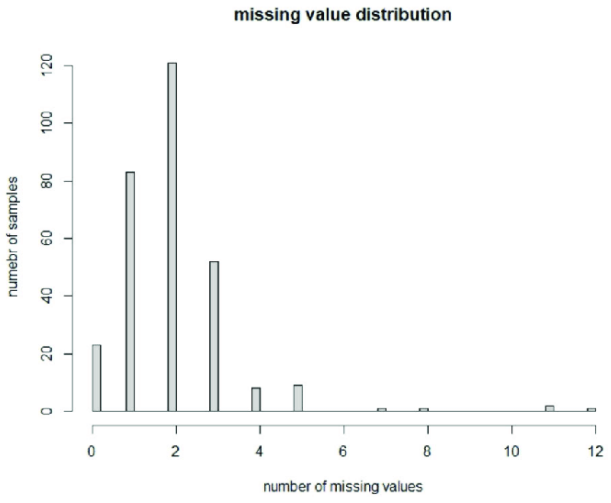
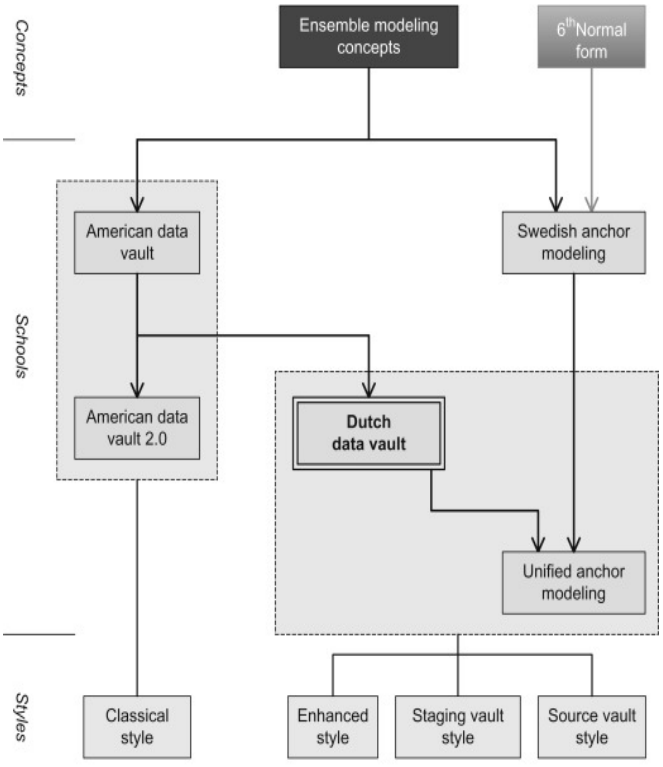


Figure 5: Histogram of missing values

Sub Curve Area (AUC): AUC measures the quality ofmodels used for separation problems. Metric forA binary number that calculates the area of less thanthe curve of a given performance measure the value of which liesbetween 0.5 and 1.Gini-Index (GI): GI is used to compare modelswhich distribution distribution is calculatedusing the Gini-coefficient and its values are between 0 and 1.

$$GI = (2 * AUC - 1). \qquad (8)$$



Note: Schools and styles renamed and simplified somewhat for presentation clarity

Figure 6: Heat map of all the features of COVID-19 data set

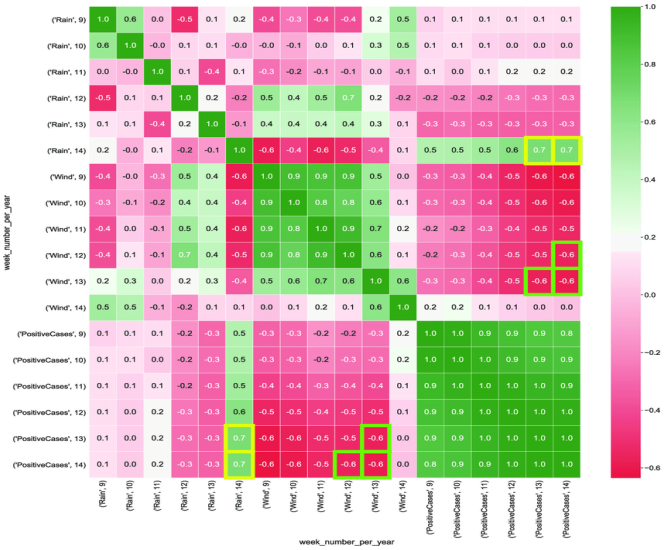


Figure 7: ensembled model.

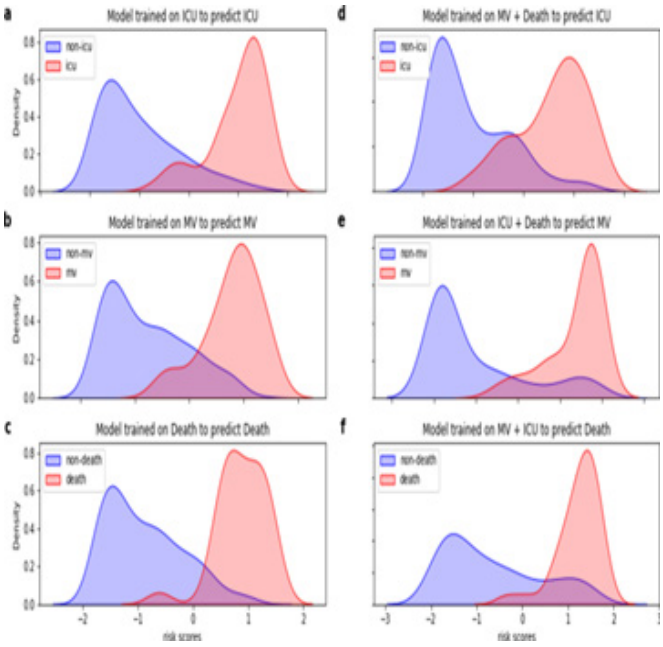


Figure 8: machine learning prediction model

KS: KS chart measures segmentation performancemodels. More precisely, K-S is a degree measureof the distinction between the distribution of good and evil.

$$KS = |\ cumulative\% + ve\ \text{-}\ cumulitttt \quad (9)$$

Error Rate (ER): ER is defined as a total rating

mis-classification count (FP + FN) is divided by a numberfor samples.

$$ER = \frac{FP + FN}{n} = \frac{FP + FN.}{FN + FP + TN + TP} \quad (10)$$

MER: Represents Minor Error Level. Herethreshold value acts as a free parameter.MWL: Related to KS statistics. Here, cost guidelineslimit value on this scale.Clarity and Sensitivity: True Positive Rate (TPR)or Sensitivity (Sens), and True Negative Rate (TNR), orcalled Specification (Spec.)

$$Sens = \frac{TP}{TP + FN}, \quad Spec. = \frac{T}{TN + FP} \quad (11)$$

Figure 7 calculates the H value using five dividers.

Typical costs are calculated on the X-axis. Allowassuming $c \in [0, 1]$ refers to the cost of not separating properlyitem of class 0 as category 1 (FP), and 1 - c stands forthe cost of improperly classifying an item of class 1 as category 0 (FN). This asymmetry may seem to emphasize KS figures,which is a simple modification of the MWL linewhere $c = \pi 1$, 1 - $c = \pi 0$. Weight rating (SR) is definedas a ratio between two costs, where SR = 1 thatrepresents symmetric costs.

$$SR = c/1\text{-}c\ ,$$

$$NormalisedCost = SR/1+SR \quad (12)$$

where, the Y-axis represents the weighted cost. TheH-measure is computed for all the five classifiers andfinally, the mean value of Severity Ratio (SR) is 1.12. Wepre-process the data to make the experimental data moreefficient and remove redundancy (Dias B.L., 2025).

Dataset

To validate the performance of the proposed CURE scheme,the dataset is being collected from the Kaggle COVID-19patient pre-condition dataset.[16] The Kaggle datasetis provided by the Johns Hopkins University throughGithub repository which contains the real-time updatedrecord of the total active cases, death cases, recoveredcases of the COVID-19 pandemic. In the modern time ofadvancement in technology and all rounded progress, tomake human beings as well as the medical science morementally and physically prepared and attentive, suchtype of health issues or threatening disease will provevery helpful and challenging. As per the reports disclosedby World Health Organization (WHO), the healthcurve (infectious cases and cured cases) remains changingabruptly every day, it becomes burdensome for themedical and other departments engaged in this kind actto serve the world medical facilities and other necessarythings to make an estimate of total requirements of thehealth related equipment's and resources. It becomesvery helpful for the entire medical department and other

concerned authorities if the corona patients be accommodatedall the resources which will prove a blessing forthem to fight the lethal disease. In this context, the datacollected contains 23 features of 5,66,603 patients.

Results and Discussion

The implementation of the experimental results are performedin Python. The results are computed based onfinding the missing values,heatmap function, featureselection, and comparison of the machine learning models.The discussion related to the results are summarized below.

Table 2: Comparison of the performance analysis of various ML prediction models.

Models	*H*	*GiniIndex*	*AUC*	*AUCH*	*KS*	*MER*	*MWL*	*Spec.Sens95*	*Sens.Spec95*	*ER*
SVM	0.687	0.802	0.901	0.901	0.802	0.099	0.098	0.443	0.447	0.46
LR	0.672	0.791	0.896	0.896	0.791	0.104	0.104	0.421	0.506	0.482
k-NN	0.655	0.781	0.891	0.891	0.781	0.109	0.109	0.478	0.49	0.469
Naïve Bayes	0.632	0.765	0.882	0.882	0.765	0.117	0.117	0.494	0.52	0.47
Random Forest	0.675	0.794	0.897	0.897	0.794	0.103	0.103	0.448	0.475	0.476

Missing Values

The initial step is to find the missing values in the Kaggledataset [16] and plot these missing values. Figure 4 visualized the histogram of the missing values in COVIDdataset. As a substitute to these, we computed the meanand replaced the missing value with its mean. The defaultinput is a numeric array with levels 0 and 1, wherethe minimum value is 0 and the maximum value is 1.

Heatmap Representation

As the Kaggle COVID-19 dataset, we collected does notcontain any missing or redundant value, so we repre. sented the complete dataset in Figure 5. It is drawn usingthe heatmap function of Python and capable to presentingthe diagrammatically view of the dataset. The parametersof the COVID patients are considered on the X andY axis.

Feature selection

As shown in Figure 6, We have selected 10 features among23 features from the COVID patient dataset. This selectionis being made by analyzing the features after computingthe feature importance score in the form of Gini-indexthrough the implementation of decision tree method.

Machine Learning Model

As discussed in the CURE scheme, the machine modelsare being used on the pre-processed data. However, there are various ways to improve the performance of speculative models based on thetechnology involved. One of those ways is to build ensemblemodels to get points for a particular result, we can begin tocombine them to produce ensemble points. Figure 7 calculates the H scale of the integrated model that can be used to improve the curve area in these models even further. Let's face it, the decision tree category and the procrastination model, both predict common risks. The new points can be calculated as a measure of these two dividers and then tested as an additional model. Usually the lower curve area develops on these combined models. After testing, the results are listed in Table 2.

Conclusion

In this paper, a CURE program is proposed based on machine learning modeling models for the treatment of COVID patients using remote e-heathcare. Performance analysis of the proposed system is tested in a Python platform that is being tested using Kaggle data from Johns Hopkins University regarding pre-patient status of COVID-19. Then, features are extracted from COVID patient data sets to diagnose symptoms-coronavirus infections. Next, the collected data was first trained and evaluated using different machine learning modeling models (such as SVM, LR, k-NN,, and Naive Bayes) that differentiate the COVID patient characteristics to predict infection rate. Finally, the performance of predictive models is tested using various metrics listed as follows: (1) H-measure, (2) Gini In-Index, (3) Area Under Curve (AUC), AUCH, KS, Minimum Error Rate (MER), Cost Minor Weight Loss (MWL), Spec.Sens95, Sens.Spec95, Error Measure (ER). Performance tests show that the CURE system exceeds the existing system relating to unequal data.In the future, we will ensure the confidentiality of corona virus data as sensitive patient information can be leaked during data transfer via wireless channels (Internet).

References

[1] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal." COVID-19 Epidemic Analysis using Machine Learning and DeepLearning Algorithms" medRxiv (2020), doi:https://doi.org/10.1101/2020.04.08.20057679.

[2] Jamshidi, M., Lalbakhsh, A., Talla, J., Peroutka, Z.,Hadjilooei, F., Lalbakhsh, P., Jamshidi, M., La Spada,L., Mirmozafari, M., Dehghani, M. andSabet, A." ArtificialIntelligence and COVID-19: Deep LearningApproaches for Diagnosis and Treatment" IEEE Access,vol. 8, pp.109581-109595,Jun. 2020.

[3] Yan, Li, Hai-Tao Zhang, Yang Xiao, Maolin Wang,Chuan Sun, Jing Liang, Shusheng Li *et al.*" Predictionof survival for severe Covid-19patients with threeclinical features: development of a machine learningbasedprognostic model with clinical data in Wuhan" medRxiv (2020).

[4] "COVID-19 Worldwide Dashboard – WHOLive World Statistics" Online available:https://covid19.who.int/, accessed on July 31,2020.

[5] Dias, B. L. (2023). Integrating Predictive Models into Public Health Policy: Forecasting Lead Exposure Risks Across the United States. International Journal of Humanities and Information Technology, 5(03), 18-38.

[6] Rehman, Suriya, Tariq Majeed, Mohammad AzamAnsari, Uzma Ali, Hussein Sabit, and Ebtesam A. Al-Suhaimi. " Current scenario of COVID19 in pediatricage group and physiology of immune and thymusresponse." Saudi Journal of Biological Sciences (2020).

[7] Nguyen, Thanh Thi." Artificial intelligence in thebattle against coronavirus (COVID-19): a survey andfuture research directions." Preprint,DOI 10 (2020).

[8] Zhang, Jian, and Yiming Yang." Robustness of regularizedlinear classification methods in text categorization." In Proceedings of the 26th annual internationalACM SIGIR conference on Research and developmentin informaion retrieval, pp. 190-197. 2003.

[9] Tan, Yuxuan." An improved KNN text classificationalgorithm based on K-medoids and rough set." In 201810th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), vol. 1,pp. 109-113. IEEE, 2018.

[10] Samuel, Jim, G. G. Ali, Md Rahman, EkEsawi, andYana Samuel." Covid-19 public sentiment insights and machine learning for tweets classification." Information,vol. 11, no. 6 Jun. (2020).

[11] Pinter, Gergo, ImreFelde, Amir Mosavi, PedramGhamisi, and Richard Gloaguen. " COVID-19 PandemicPrediction for Hungary; a Hybrid MachineLearning Approach." Mathematics, vol. 8, no. 6(2020):890.

[12] Yan, Li, Hai-Tao Zhang, Yang Xiao, Maolin Wang,Chuan Sun, Jing Liang, Shusheng Li *et al.*" Predictionof criticality in patients with severe Covid-19 infectionusing three clinical features: a machine learningbasedprognostic model with clinical data in Wuhan." MedRxiv (2020).

[13] Lin, Leesa, Rachel F. McCloud, Cabral A. Bigman,and

Kasisomayajula Viswanath." Tuning in and catchingon? Examining the relationship between pandemiccommunication and awareness and knowledgeof MERS in the USA." Journal of Public Health 39, no.2 (2017): 282-289.

[14] Hamzah, FA Binti, C. Lau, H. Nazri, D. V. Ligot,G. Lee, and C. L. Tan. " CoronaTracker: worldwideCOVID-19 outbreak data analysis andprediction." BullWorld Health Organ 1 (2020): 32.

[15] Jia, Lin, Kewen Li, Yu Jiang, and Xin Guo. " Predictionand analysis of Coronavirus Disease 2019." arXivpreprint arXiv:2003.05447 (2020).

[16] Tuli, Shreshth, Shikhar Tuli, Rakesh Tuli, and SukhpalSingh Gill." Predicting the Growth and Trend ofCOVID-19 Pandemic using Machine Learning andCloud Computing." Internet of Things (2020): 100222.

[17] "COVID-19 patient pre-condition dataset",2020. Online Available: https://www.kaggle.com/tanmoyx/covid19-patientpreconditiondataset/notebooks