A Comparative Approach To Predict Corona Virus Using Machine Learning

Rohini.M[1], Naveena.K.R[2], Jothipriya.G[3], Kameshwaran.S[4]

[1,2,3,4]Department of Computer Science and Engineering,
Sri Krishna College of Engineering and Technology,
Coimbatore, Tamilnadu, India
rohinim@skcet.ac.in, naveraj07@gmail.com,
jothipriyagurusamy00@gmail.com,
kameshkishore25@gmail.com

Jagadeeswari.M[5]
[5]Department of Electronics and Communication
Engineering, Sri Ramakrishna Engineering College
Coimbatore, Tamilnadu, India.
jagadee_raj@rediffmail.com

Abstract—Coronavirus disease (COVID-19), is one of the most infectious diseases which reshaped our everyday lives globally in the 21st century. Technology progressions have a rapid effect on every field of life, be it the medical domain or any other. More than 250 countries have been affected by COVID in no matter of time. The Indian government is making the necessary steps to control the spread of virus in the society. People all over the world are vulnerable to its consequences in the future. In a pandemic like this, people often worry whether they show a symptom of COVID-19 or not. Various AI methods have been applied successfully in epidemic studies. This paper presents the prediction and analysis of COVID-19 using various machine learning algorithms. In the present study, ML-based enhanced model is implemented to predict the possible threat of COVID-19 all over the world and the algorithms used in these models classifies the COVID patients based on several subsets of features and predicts their likeliness to get affected to this disease. This model uses 20 metrics including the patient's geographical location, travel history, health record statistics, etc., to predict the severity of the case and the feasible outcome. This research finds the patients exposed to Covid-19 and could be used as a reference, by the patients before consulting further with the doctor. The model developed using K-Nearest Neighbors (KNN) is effective with a prediction accuracy of 98.34%, Recall of 97% and an F1- Score of 0.97. Overall, this paper proposes a simple and practicable method to quickly identify and predict the high risk patients and provide priority to them for treatment so that the fatality rate can be decreased.

Keywords: COVID-19, Machine Learning algorithms, Symptoms, Classification, Prediction.

I. INTRODUCTION

The first COVID-19 case was discovered in Wuhan, China, during December 2019 and it rapidly spread in many International Countries [6]. The virus spread around the world in a very short period which became an epidemic and collapsed the health systems of many countries. Over the last few months, the virus has impacted severely with a continuous increase in the number of confirmed cases and deaths. According to WHO, globally 2.38M people have died out of 108M confirmed cases (WHO, 2021) [7]. India with the second largest population in the world stands in the top five affected countries in the world. The most likely symptoms of the virus are fever, sore throat, dry cough, headache,

instability, muscle pain, loss of taste or smell, diarrhea, and shortness of breath. The virus is easily spread among entire human community, since the pathology is highly transmitted through contact with infected person either by touch, talking, sneezing or coughing [1]. Artificial intelligence (AI) methods are used successfully to solve many problems in healthcare. AI techniques can also be used in many epidemic studies which includes the prediction of COVID-19 epidemic studies also. The overall global economy has also been affected by this pandemic along with the health, safety and hygiene of individuals all over the world. Apart from the adverse impacts of COVID-19 [8], [9], [10], there have been certain constructive influences around the world. There are no particular treatments for eliminating this disease so far, but one can minimize the COVID spread by maintaining personal sanitation and social distancing among individuals. As the world was facing losses, our nature has recovered its purity by itself from this pandemic. The harmful matters in the nature was removed from the atmosphere and the hole detected in the ozone layer was closed during this pandemic because of the complete lock down that was imposed recently. This work helps in general study of COVID-19 outbreak that utilized several machine learning classification results that visualizes the origin of disease and perform predictions and time series monitoring which helps to control the impact of disease in future. Since COVID-19 does not have very high death rate and the rate of spread of disease alone is high, when the symptoms are analyzed so far and from features, we can easily predict the important symptoms that turns disease into positive cases. Also, the healthy recovery rate implies the disease is curable yet widespread. A very important problem to be addressed is the rate of growth of infection spread by predicting the positive cases at early stage and isolating them.

Presently there is considerably less number of COVID-19 analysis kits available in hospitals which are not at all enough for the increasing cases and also due to non-awareness and fear, people undergo the test for confirming whether the result is positive or negative. Hence, it is needed to realize an automatic prediction system to effectively use the analysis kits and also to stop spreading among people by giving them proper treatment at the early stage. Machine Learning (ML) is an actually powerful tool in the fight against the COVID-19. It can be used to manage huge data and effectively predict the

spread of the disease. It helps in diagnosis and predicts COVID-19. ML Techniques are useful in tracing COVID cases, predicting, creating dashboards, diagnose and give proper medications, generating alerts to support social distance and also for other potential control mechanisms of the spread of virus.

In this paper, prediction of the symptoms of COVID is done by using various machine learning models such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RT). Section II describes the related literature works carried out in this domain. In section III, model implementation and methods are discussed. Performance metrics are explained in section IV. Experimental setup is discussed in section V and Further result analysis and performance comparison among these models is discussed in section VI and the conclusion and future scope is presented in section VII.

II. RELATED WORKS

L.J.Muhammad, et. al. [1] have proposed machine learning models for predicting COVID-19 positive cases. In this study the researcher applied deep learning network models for decision making, classification and regression tasks. The resulting classification models are used for detecting the onset of infection to human. This utilized the dataset that involves highly distributed clinical datasets including infectious and benign COVID-19 cases. This study had proved to be most accurate classification model when compared to naïve algorithms that showed 95% accuracy. Since it failed to analyze the high dimensional datasets the model implementation became complex and error prone. G.Monika and M.Bharathi Devi [2], developed three different Machine Learning models such as Polynomial Regression, Decision Tree Regressor, and Random Forest algorithm, for analyzing the COVID-19 cases. They found that the Polynomial Regression model produced the best accuracy of 90%. Still 90% accuracy is not sufficient for medical data analysis. Rajan Gupta, et. al. [3] used Machine Learning Models viz., SEIR and Regression model to analyze and predict the change in the spread of COVID-19 disease. The value of root mean squared log error for the SEIR model was found to be 1.52, and 1.75 for the Regression model. Celestine Iwendi, et. al. [4], proposed a Random Forest model boosted by the Ada Boost algorithm. The model predicts the severe ness of the positive cases. The model was able to give an accuracy of only 94% and an F1 score of 0.86 on the dataset used, which needs further increase in accuracy and F1 score. There exist a few research works regarding forecasting the disease spread but unfortunately, these studies were done by applying primitive statistic methods or by doing a simple survey. The implementation of advanced machine learning concepts in these studies is still in the formative stage.

The statistical methods and the training of ML imitating human intelligence is discussed in [15]. Several research models are identified for COVID-19 confirmed cases and in the paper [11] the reason for the spread and remedy is identified and efficient models for predicting the disorders present during COVID-19 is discussed in [12]. This research has helped the officials of the country to a certain extent for taking decisions on how to handle the situation as done in [13]. Initially, the COVID-19 Virus was mistaken as pneumonia and the analysis was done as given in [14]. The use of AI in analysis and prediction helps one to find highly accurate solutions. ML is a subgroup of AI and hence these algorithmic models are also employed in medical diagnosis and prediction [15]. All these methods and models are surveyed and finally the models suitable for this study is analyzed in this paper.

III. MODEL IMPLEMENTATION AND METHODS

This research has a sequence of steps that addresses both the infection spread and disease prediction. Initially the proper data set with necessary parameters needed for analysis must be selected and then preprocessed in the format in such a way that they can be directly applied to machine learning algorithms for processing the dataset. To implement the classifier, setting the percentage of training and test data must be done first. This is done by using trial and error method. Then each of the classification algorithms are trained and then tested. The final step is to compare the different performance metrics discussed as in section IV for all the machine learning algorithms employed in this study. The steps followed for implementation of the process is shown in Fig.1.

A. Dataset

The dataset used for analysis and prediction of COVID-19 in this paper is taken from www.kaggle.com, which is the public data set. The dataset contains 20 features as symptoms and one label with 11435 records. The various features of COVID also include Respiratory Syndromes, which provide pathological symptoms that are the features in predicting the corona virus [3]. These features lead Kaggle to the large utilized database for researchers and clinicians worldwide. Hence in this work Kaggle data set is preferred compared to other publicly available data sets. It provides several features of the disease in various stages, region-wise pathological details, the spread of disease, and the mortality rate. These are utilized in the present work to learn about infections and disease progression.
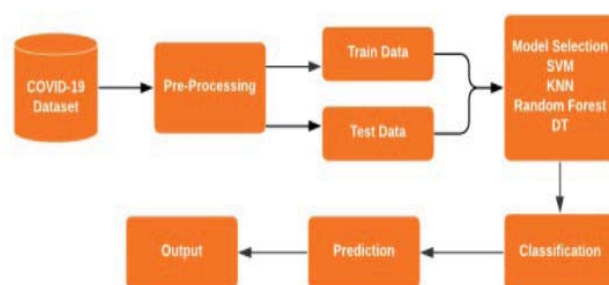

Fig.1 Process Flow

B. Preprocessing

The dataset contains categorical features, based on disease symptoms. Utilizing all the input features including significant and insignificant symptoms, training is performed initially. These features also include higher dimensional feature vectors such as confirmed, death, active and recovered COVID cases age-wise, that specifies the vulnerability of disease the excess data are pre-processed. In preprocessing, we need to convert each text category to numbers in order for the machine to process those information using mathematical equations. Clinical observations and reports are also considered for statistical examination of data and provide information about disease classification. Hence excluding non-significant and excess features and reducing the dimensionality of the derivatives provided us with more accurate performance. In this paper, we use label encoder preprocessing technique to convert the categorical feature into numeric values. This proper dataset is split-up into training and testing datasets. Training data is used to train the machine. The testing data is used to predict the label class. Training process contains both feature and label.

C. Model Selection

In this study, we employ various supervised machine learning classification algorithms to analyze and predict the positive COVID cases based on the symptoms provided by the users. SVM, KNN, DT and RF supervised learning models are applied for learning the distribution of data in the present feature selection. Choosing K-nearest neighbor technique compared to the algorithms used by the earlier researchers [2], [3] and [4], gives many calculations and distributed learning estimations. They give each model the prediction result that enable us to illustratively connect each new control group 'C' from 'N' set of features and 'K' distant vector model that gives disease classifications. To provide better significant in 'K' distance vectors the closest neighbors of feature are also analyzed and is considered for classification and prediction of infection prone individual or affected individual or normal healthy individuals. These are evaluated for performances considering the quantity of COVID features with the closest neighbors. Following are the implementation models used in the present study:

1) Support Vector Machine

Support Vector Machines (SVM) provides a set of supervised learning models which enables our present study to perform classification and regression tasks that gives accurate prediction with COVID features. Support vectors have the information points that lie closest to the decision surface which are also called as hyper planes. The optimal hyper plane is obtained from the function class with the lowest of independent features /parameters. The hyper planes are shown in Fig.2.

The points on the hyper planes, w•xi-b=+1 and w•xi-b=−1 gives ideas regarding the prediction analysis. The distance from a point (x0, y0) to a line represented by Px+Qy+=0 is given by the equation 1.

$$\frac{|Px0 + Qy0 + c|}{sqrt(P^2 + Q^{2)}} \tag{1}$$

Hence the distance between the centre axis and upper axis is calculated using the equation 2.

$$\frac{|w * x + b|}{\|w\|} = \frac{1}{\|w\|} \tag{2}$$

The total distance between the two hyper planes is calculated as 2/‖w‖. To maximize the margin, we need to minimize ‖w‖. When there exists no data points between the hyper planes, then w•xi+b ≥ +1 for upper hyper plane and w•xi+b ≤−1 for lower hyper plane.
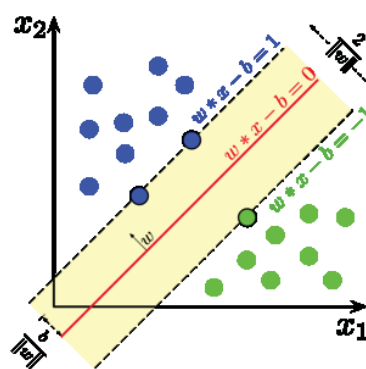

Fig.2 Maximum-margin hyper plane

2) K-NN Algorithm

Nearest neighbor algorithms are among the "simplest" supervised machine learning algorithms and have been well studied in the field of pattern recognition over the last century. While neural networks are gaining popularity in the computer vision, object detection and pattern recognition field, one area where k-Nearest Neighbor's (KNNs) models are still commonly and successfully being used for classification tasks that employ various subset of features in the computer vision. It also handles biometric data and derives patterns for classifications. KNN is an algorithm for supervised learning that simply stores the labeled training example during the training phase. KNN is also called "a lazy learning algorithm" for this reason but with highest prediction accuracy.

The process flow for KNN Algorithm is as follows

- Here we take the training sample as 'm' and the unknown point as 'p'. The samples for training are stored in array a [ ].

- For each training samples, the Euclidean distance is calculated by using d (a[i], p). Let it be denoted as 'K' neighbors.

- Count the number of data points in each category which are closer to the new data point. Then the

new data point is allotted to the category which has more closest neighbors.

- As shown in Fig.3. the new data point considered has 3 neighbors corresponding to category A and 2 neighbors corresponding to category B. in this study category A corresponds to COVID positive cases and category B corresponds to COVID negative cases. Hence the given data point is identified under category A. Thus the case is predicted as COVID positive.
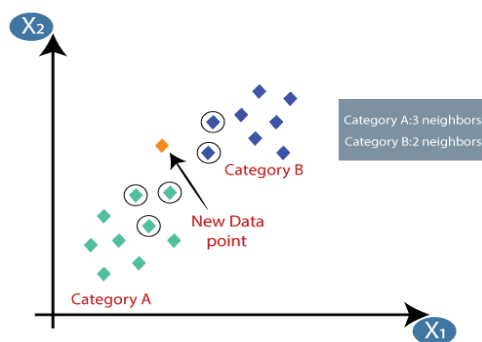

Fig. 3. KNN Classification Process

3) Decision Tree

Decision Tree (DT) is one sort of supervised learning algorithm. The entire COVID symptoms are acquired from distributed age-wise, region-wise sources and are analyzed for tree implementation. The nodes are populated until disease prediction happens for individual branch of disease symptoms till the leaf nodes. The decision tree looks like a upside-down tree with a decision rule at the root, from which subsequent decision rules are propagated to the leaf node. A decision tree has two kinds of nodes.

Step1. Every leaf node owns a class label, determined by a majority choice of training instances approaching that leaf.

Step2. Every internal node represents individual features. It branches out according to the answers (Yes or No) to the leaf node that predicts the positive cases.

4) Random Forest Classifier

Random forests involve an ensemble tree predictor in which each tree utilizes the random vector sampled with its individual features and the distribution of all trees in the classification improves the accuracy in final prediction. An eventual improvement in accuracy thus resulted from such ensemble of trees in selecting the most suitable class.

Let us consider the input features that are sampled individually for each tree in all levels from 1 to 'L' and follow the steps:

- The features in the classifier are taken as 'M' and training cases is taken as 'N'.

- To determine the decision 'm' input features is used where m<M.

- Draw a bootstrap sample 'A' that includes notable symptoms of size 'S, from the training data 'N'.

- 'm' features are chosen in random for each node of tree and the best split is calculated based on m training set variables

IV. PERFORMANCE EVALUATION METRICS

In this paper, the Confusion Matrix used, based on the actual and predicted values of four parameters with the Actual class and Predicted class parameters as shown in Table 1. The confusion matrix provides results in the matrix with total number of corona positive cases correctly as well as incorrectly predicted and corona negative cases correctly as well as incorrectly predicted by the ML models.

TABLE 1: Confusion Matrix

		Predicted Class	
		Class Yes	Class No
Actual Class	Class Yes	**True Positive**	False Negative
	Class No	False Positive	**True Negative**

TP is obtained when both actual class and predicted class of data point is 1. TN is obtained when both actual class and predicted class of data point is 0. FP is obtained when actual class of data point is 0 and predicted class of data point is 1. FN is obtained when actual class of data point is 1 and predicted class of data point is 0. Based on this, the prediction models used in this study should identify more number of TP and TN cases compared to FP and FN. To determine the performance, efficiency and quality of the four models used in this paper, four performance metrics namely accuracy, precision, recall and F1 score are evaluated.

Accuracy is the ratio of correctly predicted observations to the total observations. Accuracy is a significant factor to evaluate the system when we have distributed features where the falsely predicted as COVID positive and falsely predicted as COVID negative are almost equal. The accuracy is calculated using equation (3) and it is a great measure of performance when the used data sets are symmetrical.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad (3)$$

Precision is defined as the fraction of accurately predicted COVID-19 positive patients among all the COVID-19 positive patients. High precision relates to the low false-positive rate. The precision is calculated using equation (4). Higher value of precision metric implies that there is less number of false positive predictions. This metric gives the answer for all the patients who survived, who actually survived.

$$Pr\,ecision = \frac{TP}{TP + FP} \qquad (4)$$

Recall otherwise called as sensitivity is defined as the fraction of accurately predicted COVID-19 positive patients among all patients in the actual class. The Recall is calculated using equation (5). This metric answer to what extent the truly survived patients were labeled by the model correctly.

$$Re\,call = \frac{TP}{TP + FN} \qquad (5)$$

F1 Score is defined as the weighted average of Recall and Precision of COVID-19 patients taken into the record. The F1 score is calculated using equation (6). This metric is not very easy to understand as accuracy but it is very important and more useful than accuracy when the dataset has uneven class distribution. This metric is a useful evaluation when FP and FN predicted are equal.

$$F1Score = \frac{2*(Pr\,ecision*Re\,call)}{Pr\,ecision + Re\,call} \qquad (6)$$

V. EXPERIMENTAL SETUP AND PREDICTION MODEL

Python programming language is used to execute the algorithms in windows operating system environment. The required libraries are installed and are used for model development. To establish a strong relationship between dependent and independent features in the data set, correlation coefficient analysis is done before developing the model. The correlation of dependent and independent features is shown in Fig.4. All dependent features have positive correlation coefficient compared to independent features. The analysis has been done by using different combinations of training and testing dataset ratio. Finally the results have been plotted employing 80% of the data for training and the rest 20% for testing.

VI. RESULTS AND DISCUSSION

Predicting COVID-19 earlier can help to minimize the immense pressure in healthcare systems by which COVID-19 cases can be diagnosed which can minimize its spread among most of the nations around the world. This study has proved to give high accuracy values by applying the sci-kit-learn library. In data set, features are analyzed and pre-processed choosing best fit input that examines the disease prediction and possibility of spread that highly provides advancements in clinical treatments and decision making. All the features taken from the dataset, fits best for all the learning model discussed in this paper. Hence, we utilized the supervised learning models SVM, DT, KNN and RF for predicting COVID-19 positive and negative cases effectively. With these given

features the training is performed with labels X_{train} and given new instance X_{test} are classified. The correlation matrix of dataset features are given in Fig 4.
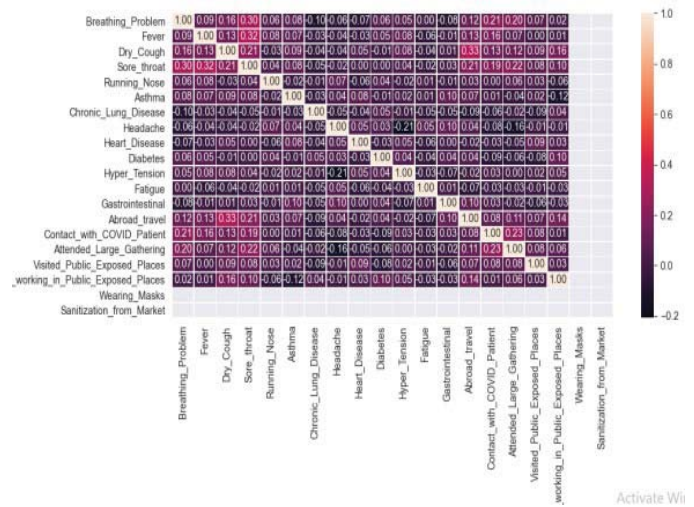

Fig 4. Correlation matrix of dataset features

The most similar memory feature to the test data that is in X_{train} is classified into a particular class. This is derived by specifying the datapoints for 'K' nearest neighbors as its parameter, providing distance metrics and significant functions on those nearest points, that are aggregated among various disease classes based on distance vectors. The performance of all the models are evaluated based on the four parameters namely accuracy, precision, recall and F1 score. The results of the performances of all the models obtained during analysis are tabulated in Table 2 and the percentage plot is shown in Fig 5.

TABLE 2: Model Evaluation

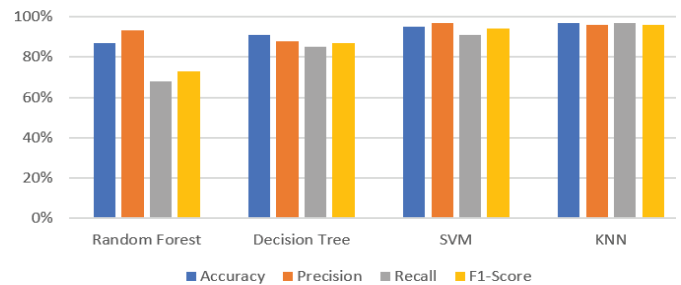| Algorithm	Accuracy	Precision	Recall	F1 Score
Random Forest	0.87	0.93	0.68	0.73
Decision Tree	0.91	0.88	0.85	0.87
SVM	0.96	**0.97**	0.91	0.94
KNN	**0.98**	0.96	**0.97**	**0.97**

Fig.5 Performance Evaluation of Developed Models

The experimental results shows that the model developed using KNN algorithm provides best performance in terms of accuracy with 98.34% when compared to other models developed with SVM, DT and RF with accuracy of 96%, 91%

and 87% respectively. In terms of precision, SVM emerges as the best model with 97% followed by KNN with 96%, RF with 93% and decision tree with 88%. KNN out performs in recall parameter among the entire model with 97% followed by SVM with 91%, DT with 85% and RF with 68%. Considering the false positive and false negative record score KNN outperforms other models in F1 score with 0.97 which is then followed by SVM with 0.94, DT with 0.87 and RF with 0.73.

The Receiver Operator Characteristic is a graphical plot which indicates the performance of binary classification problems and the Area Under the Curve measures the ability of the classifier to differentiate between positive and negative predicted classes. When AUC = 1, the classifier perfectly differentiates between all positive and negative classes perfectly. However, when AUC=0, the classifier predicts all negatives as positives and vice versa. When 0.5<AUC<1, there is a high chance for the classifier to determine the positive class from negative classes i.e., the classifier is able to detect more true positive and true negative cases than false negative and false positive cases. The ROC curve for KNN classifier with threshold value of 0.5 is shown in Fig 6.
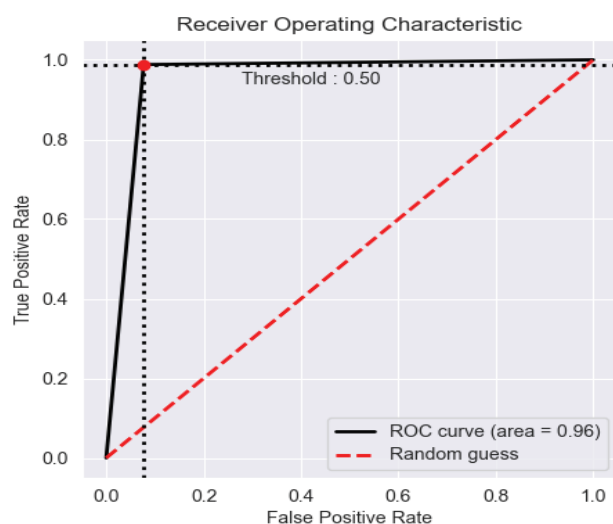

Fig.6 ROC Curve

The ROC curve shows the trade-off between sensitivity or True Positive Rate (TPR) and False Positive Rate (FPR) or 1-Specificity. The classifiers that give curves closer to the top-left corner indicate a better performance. The diagonal line corresponds to TPR equal to FPR. If the ROC curve comes closer to the diagonal then the test is less accurate. In the proposed KNN model the Area Under the ROC curve is 0.96 which means that the randomly chosen positive instances are ranked higher than the randomly chosen negative instances. The threshold value of 0.5 was identified on trial and error method and was set the same for all the models. The choice of the threshold depends on the ability of the algorithm to balance between false positive and false negative values. A confusion matrix for KNN classifier is obtained which describes the performance of the classifier as shown in Fig 7. Each row of the matrix corresponds to actual class and each

column corresponds to predicted class values. The counts of correct and incorrect classifications of COVID-19 positive and negative cases are filled in the confusion matrix. According to Fig 7, TP value of 360 means 360 COVID positive people are correctly predicted as COVID positive by the model and TN value of 1494 means 1494 COVID negative people are correctly predicted as COVID negative. 30 COVID positive cases are predicted as COVID negative and 18 Negative cases are predicted as Positive by the model for the 20% test data set provided.

Fig.8 states that the errors are distributed against all the 'K' Value of KNN algorithm, and hence we conclude that this model can be statistically accepted and therefore, can be used for prediction. The accuracy of present prediction model can be further improved by using more sophisticated exploratory data analysis and pre-processing algorithms with larger list of COVID symptoms considering the higher dimensions also.
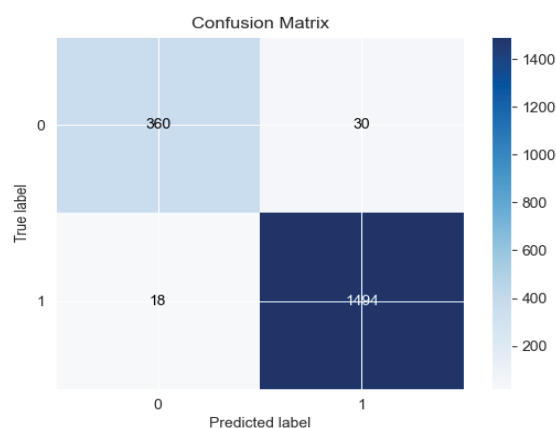

Fig.7 Confusion Matrix

VII. CONCLUSION AND FUTURE SCOPE

COVID-19 pandemic now appears to be serious infected spread disease like any other wide-spread diseases. Because of the rapid rise in the number of cases during pandemic, causes struggle in healthcare sector to identify suitable and appropriate treatment. Machine Learning methodologies are commonly used as alternative methods for classification and prediction. In this paper, Supervised Machine Learning classification models for COVID-19 virus was developed using KNN, SVM, Decision Tree and Random Forest models. All the above mentioned models are modeled using 80% data for training and 20% data for testing. The machine learning model developed with K-Nearest Neighbor performs better among all models employed for the COVID-19 data set analysis, in terms of accuracy of 98.34%, recall with 97% and F1 score with 0.97. In contrast SVM emerged to be the best model among all the models in term of precision with 97% and has less performance in other metrics compared to KNN. This study helps us to understand and stop the serious wide spread situation by utilizing the most common features that directly affects the spread of disease. This also serves the purpose of humans in predicting the disease that impacts the disease spread abnormalities and high recovery ratio.
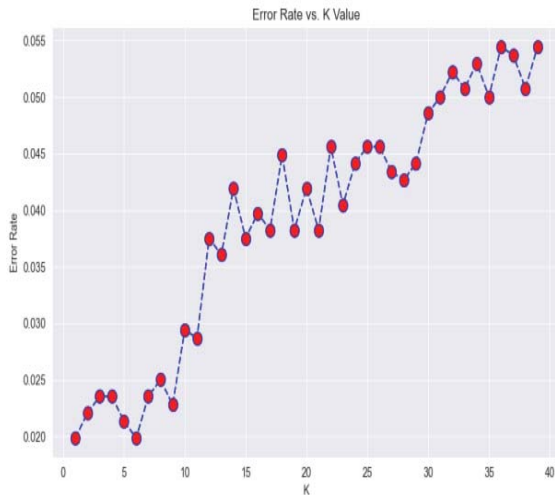
Fig.8 Error rate for KNN

The future work is to provide better significant results using several other machine learning models for finding estimates that helps the clinicians, medical and governmental organizations to look forward for real time preparations ahead at this sort of pandemic disease in future. Also Deep Learning models can be employed to predict the COVID cases.

REFERENCES

[1] L. J. Muhammad, Ebrahem A. Algehyne, Sani Sharif Usman, Abdulkadir Ahmad, Chinmay Chakraborty, I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset" Computer Science, 2020.

[2] G.Monika, Dr. M.Bharathi Devi, Using Machine Learning Approach to Predict Covid-19 Progress, International Journal for Modern Trends in Science and Technology, 6(8S): 58-62, 2020.

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

[4] Celestine Iwendi, Ali Kashif Bashir, Atharva Peshkar, R. Sujatha, Jyotir Moy Chatterjee, Swetha Pasupuleti, Rishita Mishra, Sofia Pillai, Ohyun Jo, COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm, Front. Public Health 2020. https://doi.org/10.3389/fpubh.2020.00357.

[5] S. M. Kissler et al., "Projecting the transmission dynamics of SARSCoV-2 through the post pandemic period," Science, vol. 368, no. 6493, pp. 860–868, 2020.

[6] Q. Lin. S. Zhao, D. Gao, et al., "A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action," Vol. 93, p211–216, March 4, 2020 .

[7] World Health Organization, Coronavirus disease 2019 (COVID-19) Situation Report – 42 [Internet]. WHO; 2020 [updated 2020 March 02; cited 2020 March 15].

[8] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, R. Di Napoli, "Features, Evaluation and Treatment Coronavirus (COVID-19)", Stat Pearls, 8 March 2020.

[9] Abu Al-Qumboz, M. N., & Abu-Naser, S. S, "Spinach Expert System: Diseases and Symptoms," International Journal of Academic Information Systems Research (IJAISR), 3(3), 16-22, 2020.

[10] Ajit Kumar Pasayat, Satya Narayan Pati, Aashirbad Maharana., "Predicting the COVID-19 positive cases in India with concern to Lockdown by using Mathematical and Machine Learning based Models", med Rxiv preprint doi: https://doi.org/10.1101/2020.05.16.20104133, 2020.

[11] F.Petropoulos and S. Makridakis (2020), "Forecasting the novel coronavirus covid-19," Plosone, vol. 15, no. 3.

[12] G. Grasselli, A. Pesenti, and M. Cecconi (2020), "Critical care utilization for the covid-19 outbreak in lombardy, italy: early experience and forecast during an emergency response," doi:10.1001/jama.2020.4031.

[13] www.who.int/emergencies/diseases/novelcoronavirus-2019 / technical-guidance / naming the corona virus disease (covid-2019) and the virus that causes it.

[14] "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19) in china," Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention DOI: 10.3760/cma.j.issn.0254-6450.2020.02.003 , vol. 41, no. 2, p. 145, 2020.

[15] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PloS one, vol. 13, no. 3, 2018.