Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Machine Learning based Model to combat Covid19

[1]Deepak Painuli, [2]Divya Mishra, [3]Dr. Suyash Bhardwaj and [4]Dr.Mayank Aggarwal

[1,3,4] Department of Computer Science and Engineering,

Gurukul Kangri Vishwavidhyalaya, Haridwar (Uttarakhand) 249404, INDIA

[2] Department of Computer Science and Engineering,

Uttarakhand Technical University, Dehradun (Uttarakhand) 248007, INDIA

E-mail: [1]deepak.painuli@gmail.com, [2]divya19aug@gmail.com, [3]suyash.bhardwaj@gkv.ac.in, [4]mayank@gkv.ac.in *

ABSTRACT

As declared by World Health Organization (WHO) more than nine lakh confirmed cases and more than forty six thousand death worldwide occurred due to novel coronavirus-COVID 19 from December 1, 2019 to April 1, 2020. The origin of this virus was Wuhan, China. COVID-19 has now spread all over the world and declared as pandemic disease by World Health Organization. In this paper, a dataset of 119 patients is prepared and different machine learning classification algorithm like linear classifier, K-neighbor classifier, Support Vector Machine, Decision Tree, Boosted Tree, Random Forest and Neural Network has applied to find the most suitable method that can predict the possibility of coronavirus infection. After the survey of all the algorithms it was found that Extra Tree Classifier gives best result, which is used further to predict the status of the patient. Extra Tree Classifier gives 94% accuracy. User will give input and the algorithm will predict if the patient is infected by coronavirus or not. Researchers for more accurate result can further use the method.

Keywords: *COVID-19, Pandemic Disease, World Health Organization, Machine Learning, Classifiers, Linear Classifier, Decision Tree, Neural Network, Extra Tree Classifier*

1. INTRODUCTION

Coronavirus, commonly known as COVID 19, affected 203 countries and territories around the world. It is a respiratory disease and can directly affects the immune system of the body. As stated by World Health Organization (WHO) a deadly virus called COVID 19 came into existence in the month of December 2019 in Wuhan, China. WHO declared it as a pandemic disease on March 11, 2020. From December 2019 to April 2020 the cases were increased so rapidly that it affected the life of each and every person. [1-3]

Coronavirus named as COVID-19, its main cause is cold which is similar to Middle East Respiratory Syndrome and Severe Acute Respiratory Syndrome (MERS-CoV and SARS-CoV). After comparing with these viruses, this new virus was named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV 2) disease and later COVID-19. People those have any medical history like-breathing problem, diabetes, cancer, cardiovascular disease and age above 60 will be at higher probability of infection by COVID 19. [2-5]

There are many features of COVID 19, which were target by drugs so that it is stopped in entering the cells. [6]

From recent data released by WHO it is cleared that males have higher probability of getting infected by this virus as more than 60 percent of males smoke and smoking makes the immunity system week as it makes respiratory system week due to which they might get infected by the virus.[7-9]

WHO and other health organizations have reported that SARS-CoV and MERS-CoV primarily spread from human to human. After the treatment of MERS patients, COVID 19 spread as a pandemic disease and in many countries; it takes the form of community spread in which one person can infect person's in-group, which is very dangerous phase. China has now controlled the spreading of virus by successfully implementing lockdown which other countries have to follow otherwise this virus will turn into incurable disease.[8,9]

Machine Learning (ML) is one of the key technology used in the development of process automation. Machine Learning helps doctors in such a way that they can easily take care of their patients and automate their routine processes. ML experts are working hard to develop solutions, which can diagnose and treat the disease. Machine Learning (ML) can be used to predict COVID-19 in its early stage so that its dangerous stages can be avoided and patients can be recovered easily and new patients can be minimized. [10]

Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

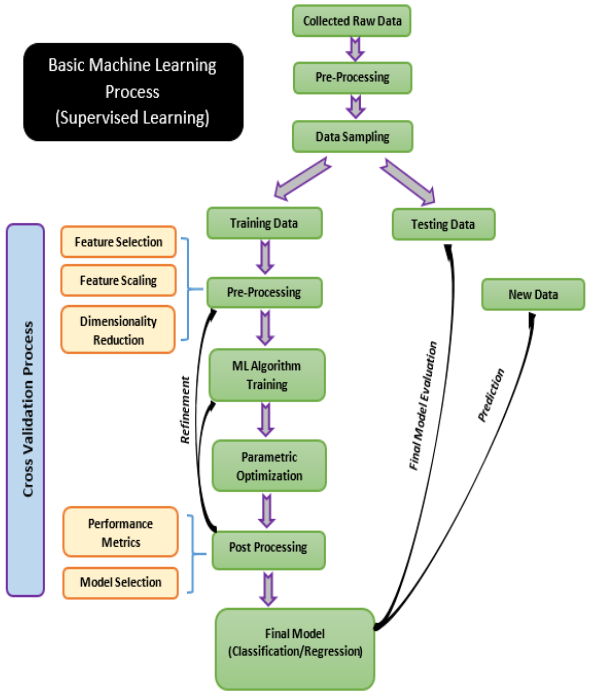

Figure 1: Flow-chart (Machine Learning Process)

Figure 1 shows the flowchart of general supervised machine learning process. This diagram shows how data is processed, cleaned in order to be able to make it ready for any particular machine learning task either it will be classification or regression. Preprocessed data is divide into Train Set and Test set for training and validation purpose. Model once trained and validated with maximum accuracy predict or classify output feature against new given input feature or symptoms.

Researchers and scientists are using many algorithms and techniques to detect or predict COVID-19. In this paper many machine learning classifiers are evaluated, to find out the best classifier in terms of accuracy, that will be applied for the prediction of COVID-19.[10, 11, 12]

The objective of the paper is to select best machine learning classifier and then train ML model using selected classifier and finally use it for COVID-19 prediction task after appropriate validation process.

.

The paper is divide in 7 sections. Section 2 gives the introduction of proposed symptom based machine learning model for prediction of COVID-19 infection. Section 3 gives brief overview of research design. Section 4 contain discussion section. Section 5 consist of performance evaluation process of ML classifiers using different-different key performance indicators (KPI). Section 6 presents results of our evaluation process and 7 focuses on conclusion and future work.

2. PROPOSED MACHINE LEARNING (ML) SYSTEM

In this paper Machine Learning(ML) classifiers are used to predict COVID-19 based on symptoms declared by WHO .Rules were prepared as per symptoms like fever ,flu, cold, age, sex, , dry cough, respiratory problem, medical history like diabetes, Cancer, breathing problem etc. , travel history, lost sense of smell known as Anosmia, and Loss of hearing ability. This system is used to predict the disease by Machine Learning classifiers-Decision Tree, Linear classifier, K-neighbor classifier, support vector machine, boosted tree, random forest and neural network taking symptoms as input feature and predict the output feature as the possibility of infected by COVID 19.[10,11]

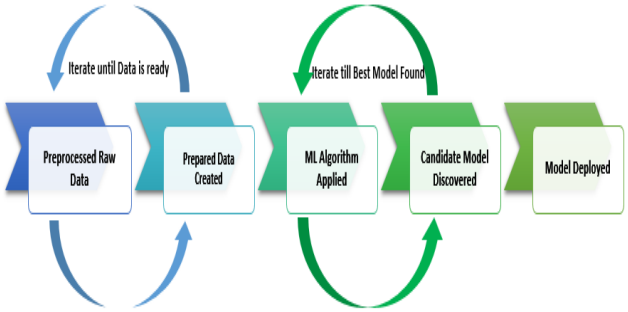

Fig. 2 Typical Machine Learning Infographic

The machine learning process applied to the rules designed as per the symptoms defined by WHO for COVID 19.Figure 2 shows the basic Machine Learning process using infographic. Machine Learning can be used for the prediction of many diseases like pneumonia, Parkinson's disease, chest disease using x-rays, arthritis disease, tuberculosis, skin disease etc.[10,11,12]

Inference rules are define, to label our target output feature "Hv-Corona" given in figure 4. If the patient is suffering from fever greater than 38^0C and has dry cough then it is the possibility that the patient might be suffering from COVID 19. If the patient above 60 years age is suffering from fever and has travelled to the country where this virus is at its peak then it is the possibility that the patient is suffering from coronavirus.[13,14]

Similarly, there are many parameters based on which these inference rules are designed like age, fever, flu, cold, sore throat, travel history, medical history, dry cough, difficulty in hearing, unable to smell and breathing problem. [2, 3]

These rules are the input variable for the classifiers to define the output variable in the form of YES or NO as this model only predicts the possibility of coronavirus infection. [14, 15].

Model is based on classification technique of supervised machine learning, where learning is supervise by labeled output feature with respect to given input features.

Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

3. RESEARCH DESIGN

Figure 3 shows the components of the system proposed in this paper for predicting COVID 19.The figure explain how this model works.

In the model, there are two phases: Training Phase and Execution Phase. In training phase, a dataset is prepared based on the details of 119 coronavirus infected patients. The details include patient symptoms like fever, cough, breathing problem, medical history, travel history etc. On this dataset number of machine learning classifiers like Linear Regression, Support Vector Machine, K-Nearest Neighbor , Decision Tree, Naïve Bayes, Random Forest, Extra Tree, AdaBoost, Gradient Boosting, Multilayer Perceptron have applied and their performance is checked in terms of F1-Score, Accuracy and AUC (Area Under Curve) and best classifier is identified.[5,16-18]

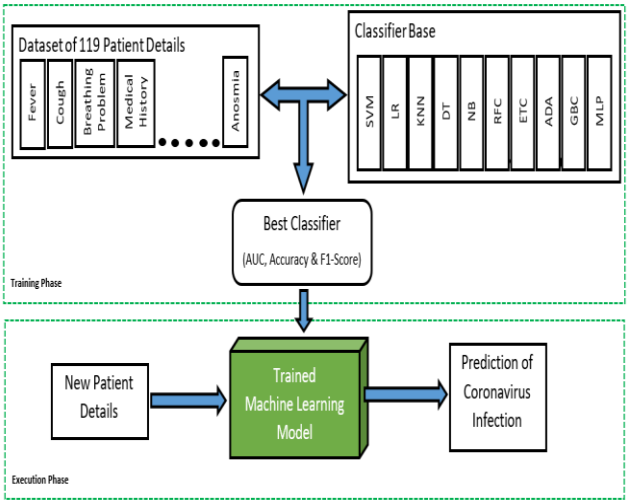

Fig. 3 Proposed Model for Coronavirus Prediction

In execution phase prediction of coronavirus has done. In this when new patient details are taken then trained ML model takes input in terms of new patient data and based on learning outcome of previously stored data of coronavirus infected patients check whether the new patient is infected by the COVID 19 or not. It predicts in terms of YES or NO i.e.: gives value 1 for YES and Value 0 for NO.

4. DISCUSSION

Classification is a term used to predict the class based on given data. Classification process is use where approximation is required from input variable, which then reach to discrete output variable. There are various algorithms available like Support Vector Machine (SVM), Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, Random forest, Extra Tree Classifier, K-Nearest Neighbor and Neural Networks (Multilayer Perceptron).etc, which can be used in medical field to predict the approximation of the occurrence of the disease. [19-20]

All these algorithms can be use with distinct features, for example, linear classifiers are used where classes are linearly separable.

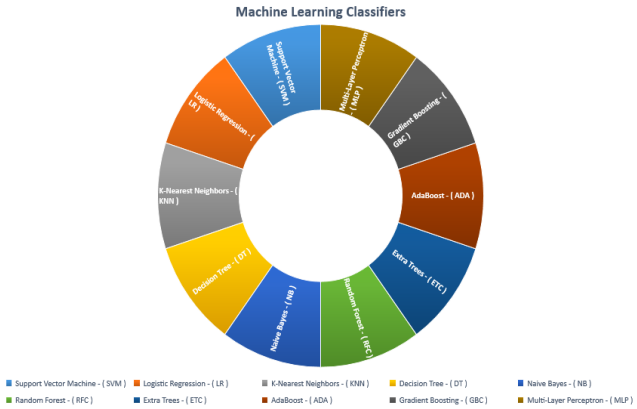

Fig. 4 Machine Learning Classifiers

Figure 4 shows the ML classifiers. Naïve Bayes comes under the category of probabilistic classifier, which is a part of Bayes Theorem in which attributes are conditionally independent. Decision Tree uses If-Then rules, which is mutually exhausted and exclusive for classification. This algorithm continuously works until termination condition occurs. Similarly, all other classifiers have their distinct characteristics. In this paper all above mentioned classifiers were evaluated to predict the probability of occurrence of COVID 19 based on symptoms specified in the dataset used and the best classifier (Extra Tree Classifier-ETC) as per the result achieved on the basis of the symptoms is suggested.[5,18,19]

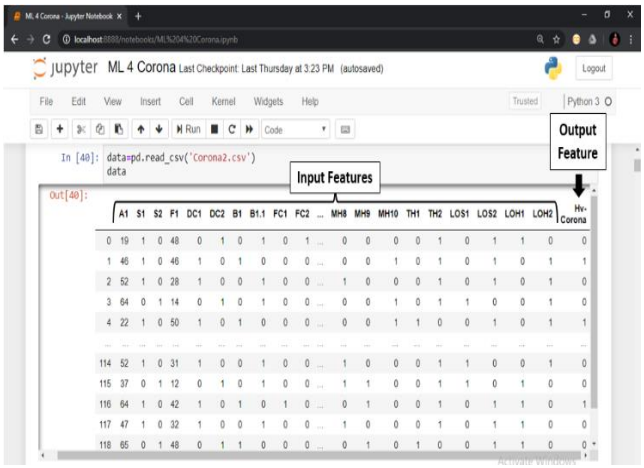

Fig. 5: Dataset of 119 Patients

Figure 5 shows the dataset of 119 patients having symptoms of coronavirus. Open source tool “Jupiter-web based IDE”, Python as programming language and its library Like Numpy, Pandas, Scikit-learn etc. are used. The symptoms have been represented in the form of symbols like A1 is used for Age, S1 is Male, S2 is Female, DC1 is Dry cough, B1 is Breathing Problem, FC is Fu and Cold, MH is medical history, TH is travelling history, LOS is Loss of

Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

smell, LOH is loss of hearing etc. and targeted output feature is "Hv-Corona". In the above dataset all values are taken in the form of 0 (NO) and 1(YES) except age (A1) and fever (F1).

5. PERFORMANCE EVALUATION

In this paper the performance of all the classifiers are measured on the basis of the input taken

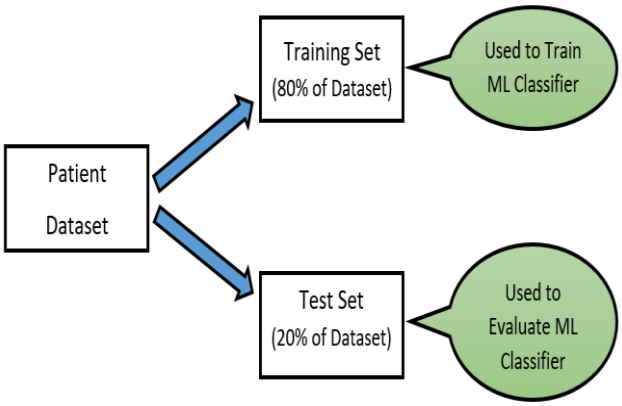

Fig. 6 Test-Train Split Process

As per the Figure 6, we split the input data into two parts; training set data and test set data. In this model the training data is taken 80% of the input and test data is taken 20% and on the basis of this data performance measurement (Area under the Curve-AUC, Accuracy & F1-Score) is done which shows which classifier is better to use for prediction purpose.

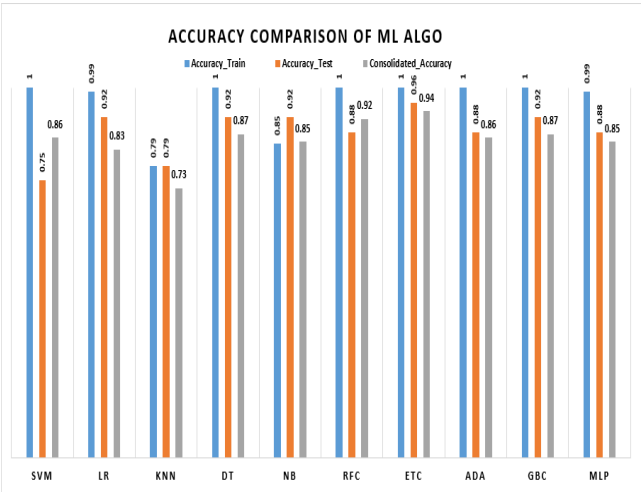

Fig. 7 Accuracy graph of ML algorithms

Figure 7 shows the accuracy comparison of different machine learning algorithms based on accuracy_test, accuracy_train and consolidated_ accuracy, which represent the accuracy shown by individual classifier over testing population, training population and combined accuracy over total population respectively. It could be observed in Figure 7 that Extra Tree Classifier (ETC) gives best result.[21,22]

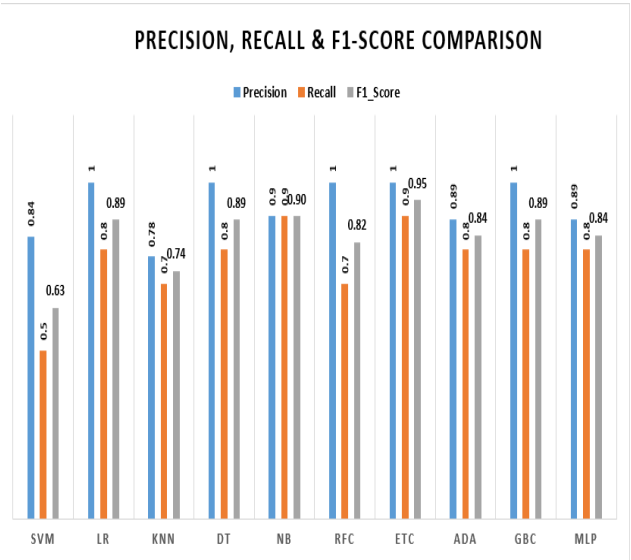

Fig. 8 Comparison Graph of F1-Score, Precision and Recall

According to the graph shown in Figure 8, comparison between three parameters – Precision, Recall and F1-Score is performed to check the performance of classifiers for performance evaluation. This allows the selection of classifiers easy as per the case.Precision is defined as the ratio of truly predicted positive observations to the total predicted positive observation, Recall is defined as the ratio of truly predicted positive observations to the all observations in actual class – "yes" and F1 score is defined as the weighted average of Recall and Precision. [23-27]

Higher the value of F1-Score, better the classifier to use. We can see as per the as per above graph Extra Tree Classifier(ETC) again shows the best result among all classifier evaluated.[23-26]

	Actual Positive	Actual Negative
Predicted Positive	True Positive(TP)	False Positive(FP) (Type 1 Error)
Predicted Negative	False Negative(FN) (Type 2 Error)	True Negative(TN)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Population}}$$

$$\text{Error Rate/Misclassification rate} = \frac{\text{False Positive} + \text{False Negative}}{\text{Total Population}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{Predicted Positive(TP+FP)}}$$

$$\text{Sensitivity/Recall} = \frac{\text{True Positive}}{\text{Actual Positive(TP+FN)}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{Actual Negative(FP+TN)}}$$

$$\text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}}$$

Fig. 9 Formulas for calculating performances classifiers

Figure 9 Shows Confusion Matrix/Contingency Table and formulas of calculating different-2 performance matrics/parameter (Accuracy-Claccification Rate, Error Rate-Missclassification Rate, Precision,Recall-Sensitivity and F1-Score) of classifiers evaluation on the basis of confusion matrix.[25]

Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Confusion Matix is a tabular layout of two dimensions ("Predicted" and "Actual") with alike set of "Classes" in both dimensions ("Positive" and "Negative"). In other words it's a table with two columns and two rows that accounts the number of true positives (TP) - correctly identified, and true negatives(TN) - correctly rejected, false positives(FP) - incorrectly identified, false negatives(FN) - incorrectly rejected. Confusion Matix permits efficient and easy interpretation of performance of classifier.[27,28]

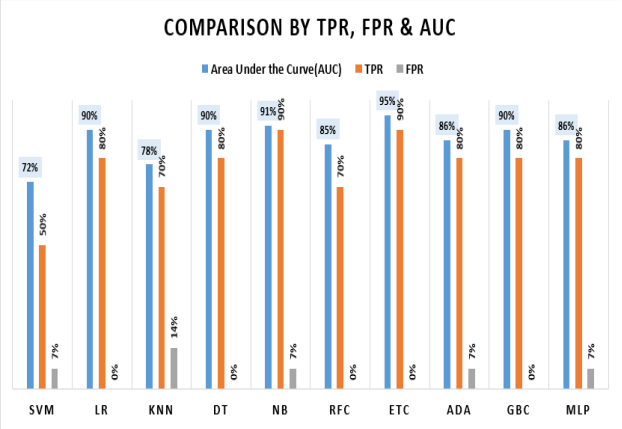

Fig. 10 Comparison Graph of TPR, FPR and AUC

Figure 10 shows the comparison of TPR, FPR and AUC. Area under the Curve (AUC) measures performance for classification problem. Higher the AUC better are the chances in predicting the difference between patients with disease and no disease. TPR is the True Positive Rate, it shows the percentage of correct positive cases that are correctly predicted by classifier and FPR is the False Positive Rate that represent the percentage of correct positive cases, which are incorrectly predicted as negative by classifier. The performance of these classifiers on the basis of TPR, FPR and AUC for the input variables prepared by the rules defined by the symptoms identified by WHO for COVID 19.[28-30]

The classifier that give high AUC and TPR and lower FPR is the best classifier in predicting the result. Therefore as per Figure 10 also Extra Tree Classifier (ETC) seems to be better classifier to use. [29, 30]

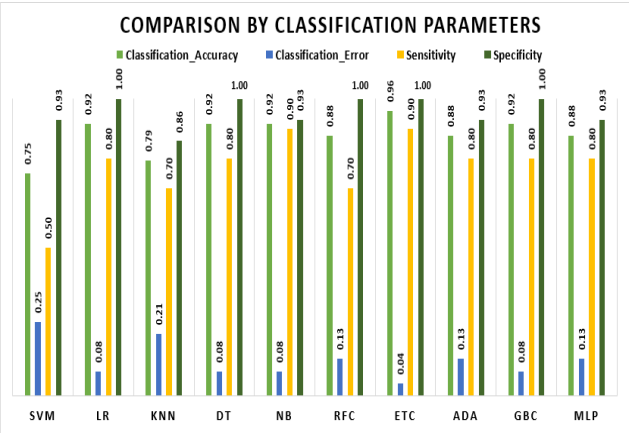

Fig. 11 Comparison by classification parameters

Figure 11 evaluates all used classifier in accuracy centric method of performance evaluation process over four dimensions(Classification_Acuuracy, Classification_Error, Sensitivity and Specificity).Classification_Acuuracy defines how accurate the classifier is.Classification_Error defines how often the classifier is incorrect or missclassification rate.Sensitivity defines if we have positive actual value then upto which extent the prediction is correct and Specificity defines if we have negative actual value then how often the prediction is correct [31-36]. As per the above graph once again Extra Tree Classifier(ETC) found to be best classifier over rest of classifier on abovesaid four parameters.

6. RESULTS

On the basis of various performance test Extra Trees classifier gives best prediction results on a given dataset (observe values in green color in Table 1) Every case requires different parameters to describe the best classifier as some requires result on the basis of Accuracy others need Precision to define output. In this model we use Area under the Curve (AUC), Accuracy and F1-Score as our prime KPI (Key performance Indicator) to evaluate ML classifiers. **Extra Trees Classifier** is a kind of entity learning technique which cluster the results of multiple de-correlated decision trees collected in a "forest" to output its classification result.

Classifier Name	Classifier Acronyms	Accuracy_Training	Accuracy_Testing	Absolute_Acc	Precision	Recall	F1_Score	TPR	FPR	Area Under the Curve(AUC)	Classification_Accuracy	Classification_Error	Sensitivity	Specificity
Support Vector Machine	SVM	1	0.8	0.86	0.8	1	0.63	50%	7%	0.72	0.75	0.25	0.50	0.93
Logistic Regression	LR	1	0.9	0.83	1	1	0.89	80%	0%	0.90	0.92	0.08	0.80	**1.00**
K-Nearest Neighbors	KNN	0.8	0.8	0.73	0.8	1	0.74	70%	14%	0.78	0.79	0.21	0.70	0.86
Decision Tree	DT	1	0.9	0.87	1	1	0.89	80%	0%	0.90	0.92	0.08	0.80	**1.00**
Naive Bayes	NB	0.9	0.9	0.85	0.9	1	0.90	90%	7%	0.91	0.92	0.08	**0.90**	0.93
Random Forest	RFC	1	0.9	0.92	1	1	0.82	70%	0%	0.85	0.88	0.13	0.70	**1.00**
Extra Trees	ETC	1	1	**0.94**	1	1	**0.95**	90%	0%	**0.95**	**0.96**	**0.04**	**0.90**	**1.00**
AdaBoost	ADA	1	0.9	0.86	0.9	1	0.84	80%	7%	0.86	0.88	0.13	0.80	0.93
Gradient Boosting	GBC	1	0.9	0.87	1	1	0.89	80%	0%	0.90	0.92	0.08	0.80	**1.00**
Multi-Layer Perceptron	MLP	1	0.9	0.85	0.9	1	0.84	80%	7%	0.86	0.88	0.13	0.80	0.93

Table. I showing the result obtained from the performance of all the classifiers

Volume 9, Issue 4
August 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

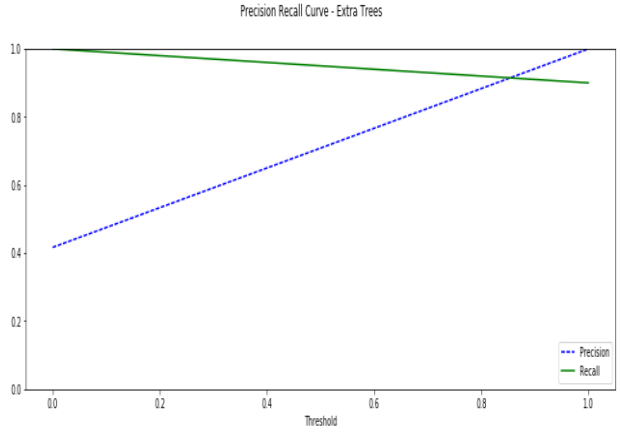

Fig. 12 Precision Recall Curve – Extra Tree Classifier

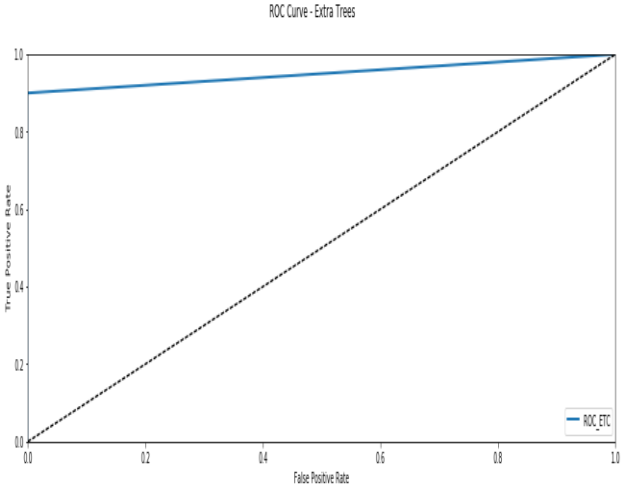

Fig. 13 ROC Curve – Extra Tree Classifier

Prediction Performance Evaluation wrt Input Features (using Test_Set)

*Entries with "Bold" & in "Orange" colour signifies wrong Pridiction by Classifier

Extra Tree Classifier - Accuracy-94%																							
Data Points(DP)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**16**	17	18	19	20	21	22	23	24
Real_Class	0	1	1	0	1	0	1	0	1	0	1	0	0	0	1	**1**	0	0	1	0	0	0	0	1
Predicted_Class	0	1	1	0	1	0	1	0	1	0	1	0	0	0	1	**0**	0	0	1	0	0	0	0	1

Data Points worngly pridicticted - 01- out of 24 [DP-16]

Fig. 14 Prediction Performance Stat using Test_Set Values– Extra Tree Classifier

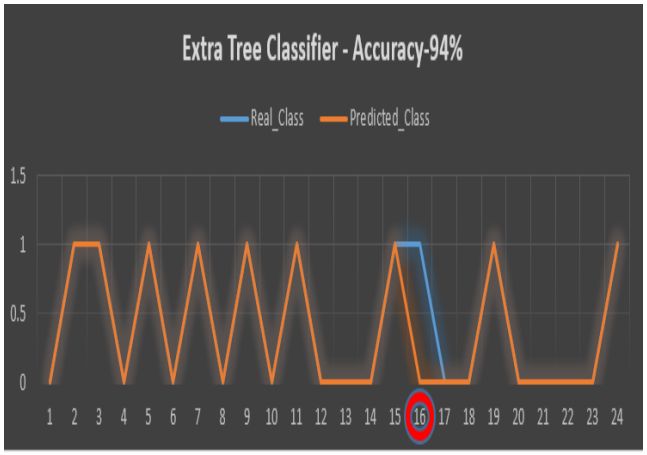

Fig. 15 Prediction Performance Graph over Test_Set Values– Extra Tree Classifier

The above figures show Precision-Recall curve (Figure 12) and ROC (Receiver Operating Characteristic) curve (Figure 13) of Best classifier as per Table 1 (Extra Tree Classifier (ETC)). ROC curve is used for the classifiers when observations of class are in roughly equal number and Precision-Recall curve is used when there is a modest to large class imbalance. Figure 14 and Figure 15 shows Prediction Performance of ETC using real class value of Test_Set and predicted class value by classifier. We can observe ETC misclassified only one data point DP-16 – (shown in "Red" color in figure 13 and encircled with "Red" color oval in Figure 14) out of 24 data points provided with Test_Set.

As it is cleared that evaluation process of ML Classifier identified Extra Tree Classifier (ETC) as best classifier to classify output feature "Hv-Corona" for the given input set, so our proposed system uses ETC as Machine Learning Model to classify user profile identified as Corona Positive or Corona Negative as per user's given feature(Symptoms) set.

7. CONCLUSIONS AND FUTURE WORK

The Hv-Corona model proved that Extra Tree Classifier (ETC) is the best classifier for the prediction of Corona Virus infection for given input dataset. The predictions done by the model can be used to save life of many people and in future it can be enhanced by taking more data.

REFERENCES

[1] Coronavirus cases. Available from: https://www.worldometers.info/coronavirus. [Accessed 18 April 2020]

[2] Fatima S.A, Hussain N, Balouch A, Rustam I, Saleem Mohd., Asif Mohd. "IoT enabled Smart Monitoring of Coronavirus empowered with Fuzzy Inference System", International Journal of Advance Research, Ideas and Innovations in Technology,2020; 06(01).

[3] Sagar Sengar discusses about How are countries around the world fighting against COVID-19. Available from: https://www.investindia.gov.in/team-india-blogs/how-are-countries-around-world-fighting-against-covid-19. [Accessed 18 April 2020]

[4] WHO discusses about Coronavirus disease (COVID-19) pandemic. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/. [Accessed 19 April 2020].

[5] Sheren Begley discusses about Who is getting sick, and how sick? A breakdown of coronavirus risk by demographic factors. Available from: https://www.statnews.com/2020/03/03/who-is-getting-sick-and-how-sick-a-breakdown-of-coronavirus-risk-by-demographic-factors/. [Accessed 19 April 2020]

[6] Michael Marshall discusses about Anatomy of a virus. Available from: https://www.newscientist.com/article/mg24532743-500-were-beginning-to-understand-the-biology-of-the-covid-19-virus/. [Accessed 18 May 2020]

[7] B. Cao, Y. Wang, D. Wen, W. Liu, Jingli Wang, G. Fan, L. Ruan ,et.al. "A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19",The new england journal of medicine, 2020; 382 (19), pp:1787-1799.

[8] Yang Z F, Zeng Z Q,Wang K et al. "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions", Journal of Thoracic Disease, 2020; 12(3), pp. 165-174.

[9] Müller M.A, Corman V.M, Jores J, Meyer B, Younan M, Liljander A, Bosch B.J, Lattwein E, Hilali M, Musa B.E, Bornstein S. "MERS coronavirus neutralizing antibodies in camels,Eastern Africa, 1983–1997", Emerging Infectious Diseases, 2014; 20(12), pp. 2093.

[10] S Sharmila , C Dharuman , P Venkatesan. "Disease Classification Using Machine Learning Algorithms - A Comparative Study", International Journal of Pure and Applied Mathematics, 2017; 114(06), pp. 1-10.

[11] Report of novel coronavirus-infected pneumonia in Wuhan City. Wuhan Municipal Health Commission, 2020.Available from: http://wjw.wuhan.gov.cn/front/web/showDetail/2020012009077.[Accessed 20 April 2020]

[12] Shirsath S.S. "Disease Prediction Using Machine Learning Over Big Data", International Journal of Innovative Research in Science, 2018; 07(06), pp. 6752-6757.

[13] Er O, Yumusak N, Temurtas F. "Chest diseases diagnosis using artificial neural networks", Expert System with Applications, 2010; 37(12), pp. 7648–7655.

[14] El-Solh A.A, Hsiao C.B, Goodnough S, Serghani J, Grant B.J. "Predicting active pulmonary tuberculosis using an artificial neural network", Chest, 1999; 116(04), pp. 968-973.

[15] Bernheim et al. "Chest CT Findings in Coronavirus Disease 2019 (COVID-19): Relationship to Duration of Infection", Radiology, June 2020; 295 (3).

[16] Tiwari U K. "Role of Machine Learning to Predict the Outbreak of Covid-19 in India", Journal of Xi'an University of Architecture & Technology,2020; 12(4),pp.2663-2669.

[17] Makridakis S, Wakefield A, Kirkham R. "Predicting medical risks and appreciating uncertainty", Foresight: The International Journal of Applied Forecasting, 2019; 52, pp. 28–35.

[18] Tuli S, Tuli S, Tuli R, Gill S S. "Predicting the Growth and Trend of COVID-19 Pandemic using Machine Learning and Cloud Computing", Internet of Things, 2020.

[19] Punn N S, Sonbhadra S K, Agarwal S. "COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms", MedRxiv , 2020.

[20] Shahid M, Shahzad Cheema M, Klenner A, Younesi E, Hofmann-Apitius M, "SVM based descriptor selection and classification of neurodegenerative disease drugs for pharmacological modeling", Molecular Informatics, 2013; 32(03), pp. 241-249.

[21] Jia L, Li K, Jian, Y, Guo X. "Prediction and analysis of Coronavirus Disease", arXiv preprint, 2020; arXiv:2003.05447.

[22] Jamgade A.C, Zade S. D. "Disease Prediction Using Machine Learning". International Research Journal of Engineering and Technology , 2019; 06(05), pp. 6937-6938.

[23] Sidath Asiri discusses about MachineLearning Classifiers. Vailablefrom: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623. [Accessed 22 April 2020]

[24] Chang W , Liu Y , Xiao Y , Yuan X, Xu X , Zhang S, Zhou S. " A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data", Diagnostics 2019; 09, pp.178.

[25] Su, Y, Shen, J,Qian, H. "Diagnosis of gastric cancer using decision tree classification of mass spectral data", Cancer Sci, 2010; 98, pp. 37–43.

[26] Austin, P.C, Tu, J.V, Ho, J.E. "Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes", J. Clin. Epidemiol, 2013; 66, pp. 398–407.

[27] DeCoste, D. "Training Invariant Support Vector Machines", Mach. Learn. 2002, 46, pp. 161–190.

[28] Parthiban G, Srivasta S.K. "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients", International Journal of Applied Information Systems, 2012; 03(07), pp. 25-30.

[29] K Thenmozhi K, P Deepika. "Heart Disease Prediction using classification with different decision tree techniques", International Journal of Engineering Research & General Science, 2014; 02(06), pp. 6-11.

[30] Na Zhu et al. "A Novel Coronavirus from Patients with Pneumonia in China, 2019", The New England Journal of Medicine , 2020; 382, pp.727-733.

[31] Kamal Alsheref F, Hasan Gomaa W, "Blood Disease Detection using Classical Machine Learning Algorithms", International Journal of Advanced Computer Science and Applications, 2019, 10(7), pp.77-81.

[32] Singh R, Singh R, "Machine Learning Based Twitter Data Mining to Analyse Sentiments of Tweets Allied to COVID-19 Epidemic & Its Patterns", ITEE Journal, 2020; 09(02), pp. 11-22.

[33] Theyazn H.H Aldhyani, Ali Saleh Alshebami, Mohammed Y. Alzahrani, "Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms", Journal of Healthcare Engineering, 2020.

[34] Matthew Mayo discusses about 7 steps of machine learning. Available from: https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html.[Accessed 25 April 2020]

[35] Mandy Sidana discusses about Intro to types of classification in machine learning. Available from: https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14.[Accessed 25 April 2020]

[36] Classification in Machine Learning .Available from: https://www.edureka.co/blog/classification-in-machine-learning/. [Accessed April 2020]

AUTHOR PROFILES

Deepak Painuli, is pursuing Ph.D from Gurukul Kangri University, Haridwar. He completed his M.Tech in 2013 and B.Tech in 2006. His research interest includes Fuzzy Logic, Machine Learning, Deep Learning, Computer Networks, and Operating System.

Divya Mishra, was born in Haridwar, Uttarakhand in 1986. Currently she is pursuing her Ph.D from Uttarakhand Technical University, Dehradun. Her research interest includes Expert System, Fuzzy Logic, Machine Learning and Deep Learning.

Dr. Suyash Bhardwaj, is working as Assistant Professor, Department of Computer Science & Engineering, at Faculty of Engineering & Technology, Gurukula Kangri Vishwavidyalaya, Haridwar. He has an experience of 7 years in academics. He has completed his B.Tech, M.Tech and Ph.D. in Computer Science with distinction. He has published many Research papers in journals of repute and has research interest in Mobile Communications, Search Engines and deep learning.

Dr.Mayank Aggarwal, is working as Associate Professor, Computer Science & Engineering at Gurukul Kangri Vishwavidyalaya, Haridwar. He has an experience of 16 years in academics. A Gold Medalist in B.Tech and completed B.Tech(Hons) in 2002. Doctorate from Gurukul Kangri in 2012 and SFRF from IIT-Delhi.