**Journal of Computer Science and Technology Studies (JCSTS)**
**www.jcsts.one**
**ISSN: 2709-104X**

# COVID-19 Classification based on Chest X-Ray Images Using Machine Learning Techniques

Muhammad Imad[1]*, Naveed Khan[2], Farhat Ullah[3], Muhammad Abul Hassan[4], Adnan Hussain[5] & Faiza [6]
[1,3,4,6]*Department of Computing and Technology, Abasyn University, Peshawar,* Pakistan
[2]*Department of Computing, Ulster University, Belfast, United* Kingdom
[5]*Department* of Computer Science, Islamia College University, Peshawar, Pakistan
**Corresponding Author:** Muhammad Imad, E-mail: imadk28@gmail.com

**Received**: August 02, 2020
**Accepted**: October 01, 2020
**Volume**: 2
**Issue**: 2

**KEYWORDS**

COVID-19, Support Vector Machine, K-Nearest Neighbors, Random Forest, Naïve Bayes algorithm, Decision Tree

**ABSTRACT**

The coronavirus (COVID-19) pandemic rapidly spread from the infected person who has a severe health problem around the world. World Health Organization (WHO) has identified the coronavirus as a global pandemic issue. The infected person has a severe respiratory issue that needs to be treated in an intensive health care unit. The detection of COVID-19 using machine learning techniques will help in healthcare system about fast recovery of patients worldwide. One of the crucial steps is to detect these pandemic diseases by predicting whether COVID19 infects the human body or not. The investigation is carried out by analyzing Chest X-ray images to diagnose the patients. In this study, we have presented a method to efficiently classify the COVID-19 infected patients and normally based on chest X-ray radiography using Machine Learning techniques. The proposed system involves pre-processing, feature extraction, and classification. The image is pre-processed to improve the contrast enhancement. The Histogram of Oriented Gradients (HOG) is used to extract the discriminant features. Finally, In the classification step, five different Machine Learning algorithms, such as Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest, Naïve Bayes algorithm, and Decision Tree) are used to efficiently classify between COVID-19 and normal chest X-ray images. The different metric measures like accuracy, precision, recall, specificity and F1are used to analyze the results. The result evaluation shows that SVM provides the highest accuracy of 96% among the other four classifiers (K-Nearest Neighbors and Random Forest achieved 92% accuracy, 90% accuracy of Naïve Bayes algorithm and 82% accuracy of Decision Tree).

## 1. Introduction

In December 2019, the coronavirus (COVID-19) pandemic was identified by Wuhan city in China, which has a severe public health problem worldwide (Yadav, 2020) (Hassanien, 2020). The cause of the coronavirus (COVID-19) is Severe Acute Respiratory Syndrome (SARS). The Hubei region in china was identified two different coronaviruses, which are Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) (Hassanien, 2020). Coronaviruses spread from cold to severe diseases that can affect both humans and animals (Wu, 2020).

The World Health Organization (WHO) identified that the total case of coronavirus (COVID-19) infected is 2,995,758 is confirmed and 204,987 deaths are reported from the coronavirus (COVID-19) pandemic outbreak till the 29 April 2020 throughout the worldwide (WHO, 2020). In many developed countries, the health system collapses due to spread COVID-19 simultaneously and to increase the demand for intensive care units filled with patients who are infected from the COVID-19[1]. The COVID-19 infected cases are increasing continuously worldwide from starting to till date (29 April 2020) is presented in Figure.1. The statistics about Total Confirmed Cases, Total Death and Total Recovered from the infected of COVID-19 (WHO, 2020) are presented in the following Figure 1.

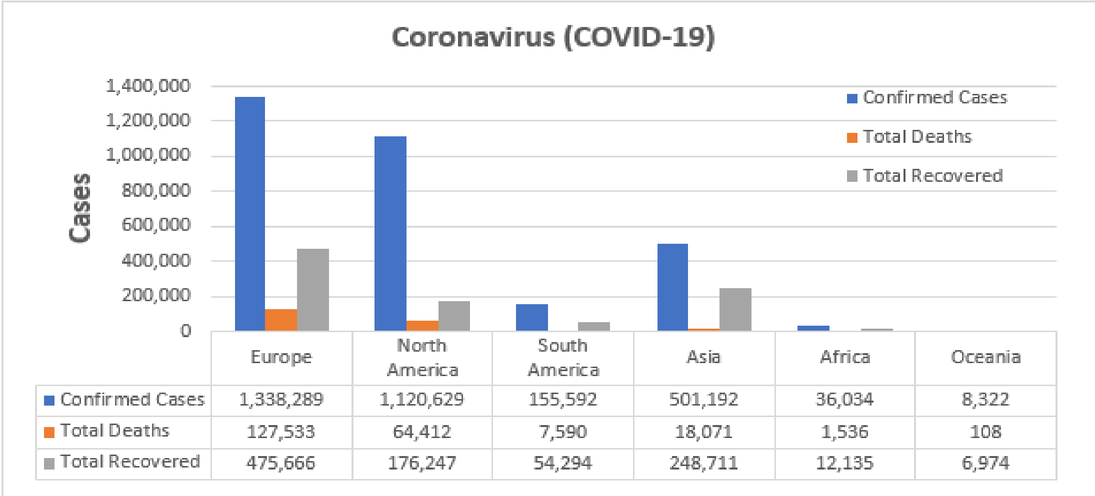



Fig. 1. COVID-19 cases worldwide (Region wise) [4]

The spread of COVID-19 disease transmission from human to human is quite fast, and increasing exponentially (Wu, 2020). The highest cases reported from outsides to china are the US, Spain, Italy, United Kingdom, Germany, France, Turkey, Russia, Iran, Brazil, Canada. According to World Health Organization (WHO), the US now has the highest number of new confirmed cases of 983,457, which is increasing rapidly, and the death ratio reached 50,492 while the new cases are 22,541 reported recently (WHO, 2020). After the strict lockdown and intensive care treatment in China, Spain, Italy, Turkey, Canada, and other countries, the rate of deaths and new cases is going to decrease, which is shown in Figure 2.

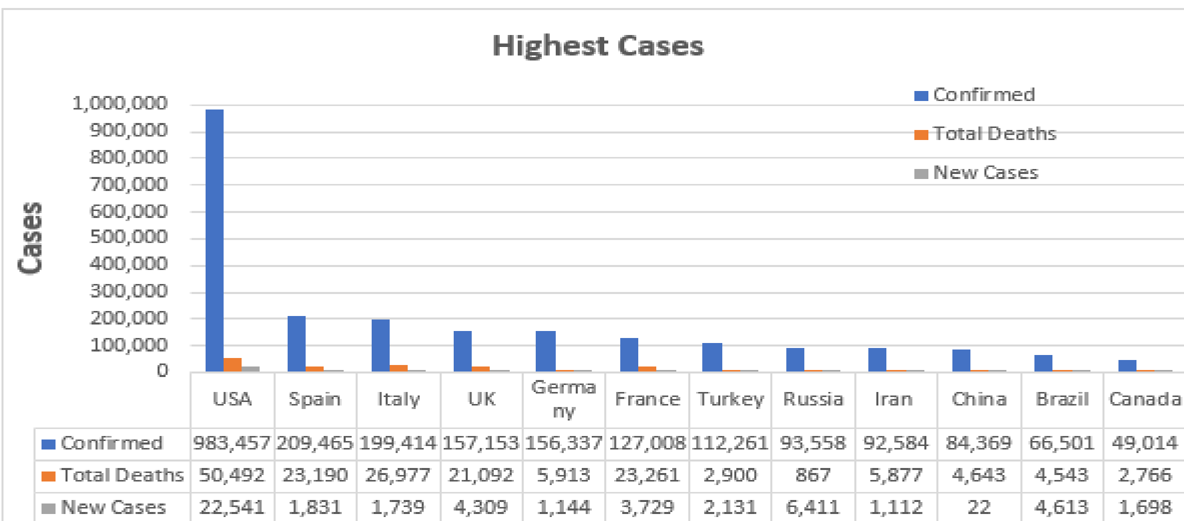



Fig. 2. COVID-19 highest cases (Country wise) [4].

The coronavirus (COVID-19) spreads from the infected person through droplets of saliva or having coughs or sneezes. Most people who are infected with coronavirus (COVID-19) are moderate in respiratory illness and having medical problems such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop severe illness. Those ages over 60 years old and underlying medical conditions are having a high risk of infected from COVID-19 disease. At this time, there are no approved vaccines or treatments for COVID-19. However, to protect yourself and others from infection of coronavirus (COVID-19) by washing hands, avoid touching face and wear a face mask, maintain at least a 1-meter distance who having coughing or sneezing problem (WHO, 2020).

The common symptoms of COVID-19 include fever, tiredness, dry cough, shortness of breath, aches and pains, sore throat, and sometimes diarrhea, nausea or runny nose. The men are more infected than women and there is no death in children between the age of 0-9. The cases of infected COVID-19 are growing quite fast worldwide due to careless nature in humans (Ozkaya, 2020).

The rest of the paper is organized as follows. Section 2 explains the related works of COVID19. Section 3 briefly explains the Proposed Methodology, such as pre-processing, feature extraction, classification steps. Results and discussion are presented in section 4 and finally, the conclusion is presented in Section 5.

## 2. Literature Review

Coronavirus starts in china and spreads rapidly among other countries. The COVID-19 can live on different surfaces for a few days like steel, plastic, copper, cardboard. The novel approach is used to diagnose and classify the COVID-19 using Support Vector Regression method to identify five different tasks related to coronavirus. The tasks are to find the spreading of COVID-19 across regions, the growth rate of COVID-19 infected patients and their mitigation, how the' epidemic's end, analyzing the transmission rate of the virus, and the correlation of COVID-19 with weather conditions. The result is compared with other regression models (Linear regression, Polynomial regression), which providing promising results both in accuracy and efficiency (Yadav, 2020).

China's province of Hubei has been identified that there are two different coronaviruses, which are Middle East respiratory syndrome and Severe acute respiratory syndrome. Therefore, the COVID19 has been detected from infected person X-ray images using multi-level thresholding with the Support vector machine. The size of all images is 512*512 and JPEG format. The accuracy achieved of the proposed model is about 97.48% (Hassanien, 2020).
The coronavirus is the fastest spreadable virus among people due to severe acute respiratory syndrome. Jiangpeng Wu presented a method to accurately identify the coronavirus from the CT x-ray and the symptoms.253 samples of the infected COVID-19 patients were collected from a different source. 49 'patients' blood was tested in the commercial clinical center, where 24 patients were infected from the COVID-19. The method accurately identifies infected patients with an accuracy of 95% and a sensitivity of 96.95% from the cross-validation method (Wu, 2020).

The most important symptom for the diagnosis of COVID-19 is cough, breath shortness and fever. A novel method has been proposed for the detection of coronavirus (COVID-19). The feature fusion and ranking method have been used to increase the performance accuracy. Further, the data has been classified by using a support vector machine (SVM). The deep feature was obtained by the pre-trained Convolutional Neural Network (CNN) model in the Subset-1 and Subset-2 datasets. The proposed model obtained a high accuracy of 95.60% in subset-1 and 98.27% in the subset2 dataset (Ozkaya, 2020).
Early-stage diagnosis of the COVID-19 can help to prevent the diseases from spreading. Abbas presents the technique to detect the COVID19 from the comprehensive image dataset of chest X-ray images. For this purpose, previously implemented CNN architecture Decompose, Transfer and Compose (DeTraC) have been used to classify the COVID-19 infected and normally based on chest X-ray images. The result analysis from the proposed model achieved a high accuracy of 95.12% from the DeTraC mechanism in classification (Abbas, 2020).

The coronavirus (COVID-19) does not have any proven drug treatment or vaccine. The machine learning techniques have been used to detect real-time patients of coronavirus (COVID-19) from the chest CT scan images. The image is pre-processed by using an adaptive winner filter to reduce the noise and improve the quality. The FFT spectrum is used for the texture feature and then classifies the image using a linear support vector machine (SVM). The FFT-Gabor scheme has shown that the prediction in real-time of the coronavirus (COVID-19) patient with an accuracy of 95.37%, sensitivity95.99%, and specificity is 94.76% (Al-Karawi, 2020).

According to WHO, the COVID-19 is a pandemic disease. The X-ray and CT images are considered for detection and infected region localization of COVID-19 from two different modalities of medical imaging. The IRRCNN and NABLA-3 network models have been applied for the classification and segmentation process. The result of the proposed algorithm shows that 84.6% and 98.78% result has been achieved from the X-ray and CT images, respectively (Alom, 2020).

The most suitable method of controlling over the coronavirus is quarantine and appropriate treatment. The Artificial Intelligence technique and Machine Learning algorithms have been used for the detection of coronavirus (COVID-19) from the chest X-ray images. The images were classified using CNN with the SoftMax classifier, SVM, and random forest. CNN is used for two scenarios; the image is classified and extract the graphical feature for a hybrid system. The extracted feature has been used further for training and testing parameters. The proposed algorithm shows that CNN accuracy is 95.2%, which is better than other methods (Alqudah, 2020).

The coronavirus infected person has severe respiratory problems that need to be treated in an intensive care unit. The death rate of COVID-19 is less than 3%, while the casualty rate of severe is high.336 cases of COVID-19 infected patients are divided into training and test dataset.220 clinical records were collected to identify the symptoms. The support vector machine classifies the age, GSH, CD3 and protein with an accuracy of 77.5% and the specificity is 78.4%. The AUROC reaches 0.9996 in training and 93.33 in testing. While, the cut-off value is 0.0667 and the recall rate is 93.333% (Sun, 2020).

A deep-running technique that is based on X-Ray images to identify the COVID19. The supported vector machine classifies the X-Ray images from deep features. The proposed classification model of the resnet50 plus and SVM achieved an accuracy of 95 %. The data set has been downloaded from the online repository of GitHub, Kaggle site (Sethy, 2020).

Detecting the coronavirus (COVID-19) based on radiography and radiology images is the most efficient way to treat the patients. A deep learning framework has been proposed for the detection of coronavirus (COVID-19) from chest X-ray images. The four pre-trained convolutional models, including ResNet18, ResNet50, Squeeze Net, and DenseNet-121, have been used to identify the coronavirus (COVID-19) disease. The two datasets are considered and combined, which have been around 5k images called COVID-Xray-5k. The result from the all pre-trained model shows that sensitivity 97% and specificity 90% is achieved (Minaee, 2020).

The only specific way to diagnose the virus by taking a test. A deep learning technique (ResNet-18, ResNet-50) has been applied to classify the coronavirus (COVID-19) from CXR images. The COVID-19 CXR datasets (CORDA, RSNA, COVID-Chest X-ray, Chest X-ray) have been collected from the emergency hospital in Northern Italy. The result analysis from the proposed deep learning algorithms shows that the training with a COVID-Net architecture and COVID-Chest Xray dataset is 85% accuracy and the ResNet-18 architecture with a COVID-Chest Xray dataset is 100% accuracy) (Tartaglione, 2020 ).

The COVID-19 pandemic is rapidly growing in all over the world. A deep neural network (DNN) related method has been proposed to identify the COVID-19 from the Chest X-Ray datasets consisting of normal, pneumonia, and COVID-19 images. The images are pre-processed to highlight the class-discriminated region using Layer-Wise Relevance Propagation (LRP) and Gradient-Guided Class Activation Maps (Grad-CAM++). The results of the proposed algorithm show that the Positive Predictive Value (PPV) achieved 89.1% accuracy and 83% recall has been predicted (Karim, 2020).

In this paper, the authors proposed a model for early diagnose of coronavirus using four image filtering techniques (MPEG7 Histogram Filter, Gabor Image Filter, Pyramid of Rotation-Invariant Local Binary Pattern Histograms Image Filter, Fuzzy 64-bin Histogram Image Filter) with hybrid feature selection model. The feature is extracted from the CT images through conventional statistical and machine learning tools. The selected feature is classified by using a stacked hybrid classification system. The result of the proposed model shows that Naïve Bayes achieved the highest accuracy of 96% as a meta classifier in a hybrid classification method (Farid, 2020).

## 3. Material and Methods

The dataset and proposed methodology (Pre-Processing, Feature Extraction and Classification) has been discussed in detail in this section.

### *3.1 Dataset Description*

In this study, different coronavirus (COVID-19) dataset has been combined, which is obtained from the open sources Kaggle repository (Kaggle, 2020).). The total images in the dataset repository are 2200 categorized by Chest X-Ray and CT Images of a COVID-19 infected (1400 images) and Normal patients (800 images). 80% of the dataset is used for training and 20% for testing purposes. The X-ray images were different in gray levels, patch size, dimensions, and characteristics were obtained. Figure 3 illustrates a sample X-ray image of a COVID-19 infected patient and a normal patient used for the research. The flow diagram of the proposed model is presented in Figure 4.

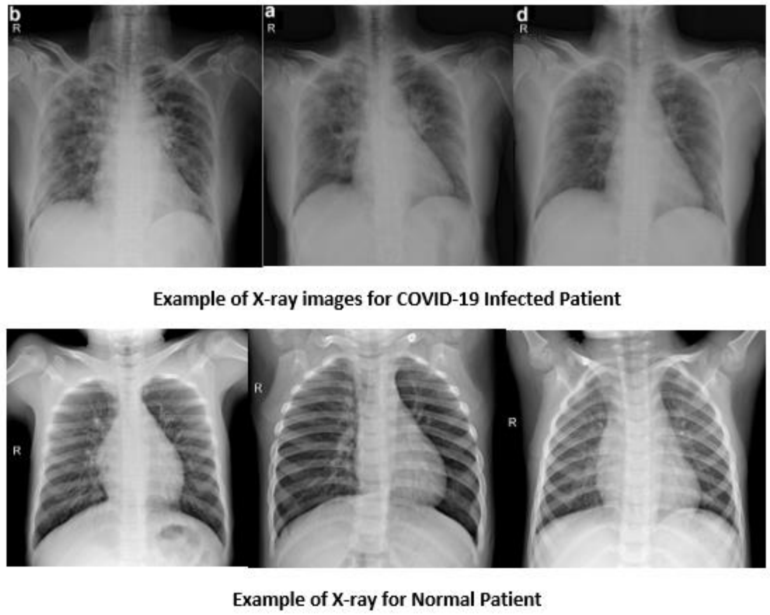



Fig. 3. Example of chest X-ray images of COVID-19 infected patients and Normal

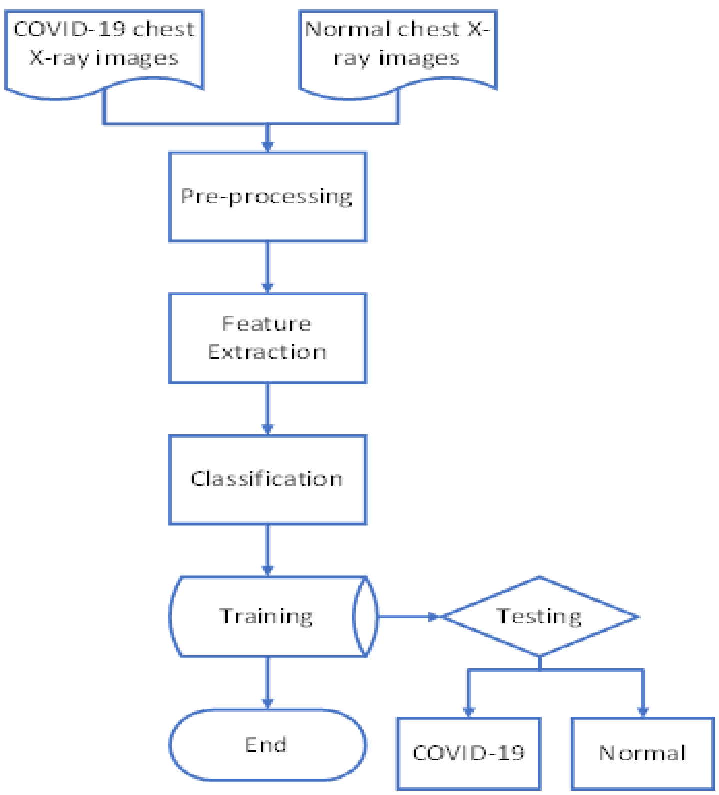



Fig. 4. flowchart of the proposed system

***3.2 Pre-processing***

The radiographic images always have dark edges and noise (Karim, 2020). The pre-processing step is necessary to improve its generalization for removing noise and improve the contrast enhancement in the entire image by applying a median filter, average filter and histogram equalization. A median filter is used as a nonlinear method for removing salt and pepper noise while preserving edges. The median filter first sorts all the pixel in the image and replace with a median of the neighborhood pixels. The average filter smooths the image by minimizing the intensity variation of neighboring and replace the average value of a neighboring pixel, including itself (Auckland, 2020). Moreover, Histogram equalization improves the contrast in an image to stretch out the intensity range, which provides a better quality without loss of any information (Wikipedia, 2020).

### *3.3 Feature extraction*

The Histogram of Oriented Gradient (HOG) is a feature descriptor technique presented in (Dalal, 2005) is used to extract the feature from COVID-19 infected patients and normal X-ray images which are separated in a positive and negative region. Different steps are involved in HOG, which is presented in (Started, 2020). These features are used to build machine learning models.

### *3.4 Classification*

After the feature extraction step, different machine learning algorithms such as Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, Decision Tree, and Random Forest are used to perform classification.

#### *3.4.1 Support Vector Machine (SVM)*

Support Vector Machines uses some data collection to supervise the learning models. SVM analyzes the data used to train and test the model for classification and regression problems. The SVM constructs a hyperplane in multidimensional space, which is separated by different classes to minimize any error. The maximum hyperplane (MMH) provides a better result when dividing the data into two classes (Barstugan, 2020) (Saeed, 2020). After feature extraction from the X-ray dataset using HOG, the SVM is used to train and test the datasets having COVID-19 infected patients and normal images.

#### *3.4.2 K-Nearest Neighbor (KNN)*

KNN algorithm is a type of supervised machine learning algorithm that is used for the discrete labels data as well as continuous labels data problems to assign the class to a new data point. KNN keeps the training data by computing the similarity between the input data and the training instance to predict the label (KNN, 2020) . The main steps in KNN are, calculate the distance and find the closest neighbor. Each object votes for the class; the most vote for their class is considered as a prediction (KNN, 2020).

#### *3.4.3 Naive Bayes*

Naïve Bayes algorithm is based on Bayes theorem which is used for classification. This technique was introduced in the 1960s under the method of text retrieval problem and was mostly used for text categorization (NB, 2020). Bayes theorem provides to calculate the posterior probability of $p(\frac{a}{b})$ from $p(a)$, p(b) and $p(\frac{b}{a})$, which is shown below in equation 1.

$$p\left(\frac{a}{b}\right) = \frac{p(b/a)p(a)}{p(b)} \qquad (1)$$

Equation 1 presents the posterior probability of p(c/x) in a given predictor where (c is target and x is the attribute). p(c) and p(x) is the prior probability of the class and predictor respectively. p(x/c) is the probability of a given class (Naïve, 2020).

#### *3.4.4 Random Forest*

Random forest is a supervised learning algorithm used for classification and regression analysis. The algorithm was created by Tin Kam ho by using random subspace method (Mandibules, 2020). The random forest consists of a large number of decision trees which operate as an ensemble. Each tree spits out for a class prediction based on the taking majority votes in tree class. The main steps for the random forest are, select the random sample from the dataset and construct a decision tree to get a prediction from each tree. Perform a vote for the prediction tree and select for the final prediction based on the most vote (RF, 2020).

#### *3.4.5 Decision Tree*

The decision tree is a non-parametric supervised learning method that is used for classification and regression analysis. It worked on continuous and categorical output variables (DT, 2020). In a decision tree, there is a two-step process for classification, learning step and prediction step. In the learning step, the model is learned from the given training data, and in the prediction step, the model is used to predict the response for the given testing data (DT, 2020).

### *3.5 Performance Evaluation*

The confusion matrix is most commonly used in machine learning to evaluate the performance of the classification model. The number of correct and incorrect result in a classification problem are summarized and compare the output with the reference data. The confusion matrix specifies the most common matric, such as Accuracy, Precision, Recall, Specificity, and F1-score. To solve the confusion matrix, four statistical indices have been used that is a true positive (TP), true negative (TN), false positive (FP), false negative (FN) were calculated, which is given in equation (2) to (5).
True Positive (TP) refers to predicted yes that they have infected by COVID-19, True Negative(TN) refer to predicted no, that they don't have infected by COVID-19, False Positive(FP) refer to predicted yes, but they don't have infected by COVID-19, False Negative (FN) refer to predicted yes, but they have actually infected by COVID-19.

$$\text{Accuracy} = \frac{(TN+TP)}{TN+TP+FN+FP} \quad (2)$$

$$\text{Precision} = \frac{(TP)}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{(TP)}{TP+FN} \quad (4)$$

$$\text{F1-Score} = 2 * \left(\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}\right) \quad (5)$$

From the above equation (2) to (5), accuracy indicated that how often the classification is correct, while precision indicates how often it will correct classify during prediction. However, recall indicates that from the all positive class how much predicted correctly while specificity evaluates that the negatives that are identified correctly. The F1-score indicates the average of precision and recall (Confusion matrix, 2020).

**4. Results and Discussion**

In this study, a chest X-ray dataset has been used for the prediction of coronavirus (COVID-19) infected patients and normal). The different machine algorithms discussed earlier are implemented in python using a Dell laptop with corei5,8GB of RAM to perform experiments. The dataset has been trained and tested on chest X-Ray images using machine learning techniques such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, Decision Tree, and Random Forest. The results of the different machine algorithms are presented in Table 1. The results show that the highest accuracy was obtained using the SVM model. The K-nearest neighbor and Random forest have a similar performance. More ever, the SVM algorithm achieved highest accuracy of about 96%, and precision is 95%, recall of 92, specificity 92% and F1 is 96%. However, the decision tree achieved the lowest performance in terms of accuracy (82%), precision (86%), recall (87%), specificity (73%) and F1 score (86%). The results in Table 1 presents that the support vector machine achieved the highest prediction results using different metric measures such as accuracy, precision, recall, specificity and F1 score as compared to other machine learning models.

**Table 1.** Performance results are obtained from different Machine Learning algorithm.

| Classifications | Confusion Matrix | | | | Performance Result % | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | TP | FP | TN | FN | Accuracy | Precision | Recall | Specificity | F1 |
| Support Vector Machine (SVM) | 267 | 13 | 145 | 5 | 96% | 95% | 98% | 92% | 96% |
| Decision Tree | 248 | 39 | 107 | 36 | 82% | 86% | 87% | 73% | 86% |
| Naive Bayes | 236 | 34 | 153 | 7 | 90% | 87% | 97% | 81% | 92% |
| k-Nearest Neighbors (KNN) | 248 | 22 | 150 | 10 | 92% | 92% | 96% | 87% | 94% |
| Random Forest | 279 | 5 | 117 | 29 | 92% | 98% | 90% | 95% | 94% |

In Figure 5, the Confusion matrix was obtained for COVID-19 infected patients and Normal patients of the proposed machine learning techniques are given. When the confusion matrix was evaluated, the Support vector machine is classified with COVID-

19 infected patients of 267 images as true positive and normal images of 139 as true negative, which achieved the highest accuracy result of 96%. The KNN and Random Forest have an accuracy of 92%, while the KNN true positive images are 248 of COVID-19 and 150 as a true negative image. The lowest performance accuracy is achieved by a decision tree from 248 images as a true positive of COVID-19 and 107 as a true negative normal image, which is illustrated in figure 5.

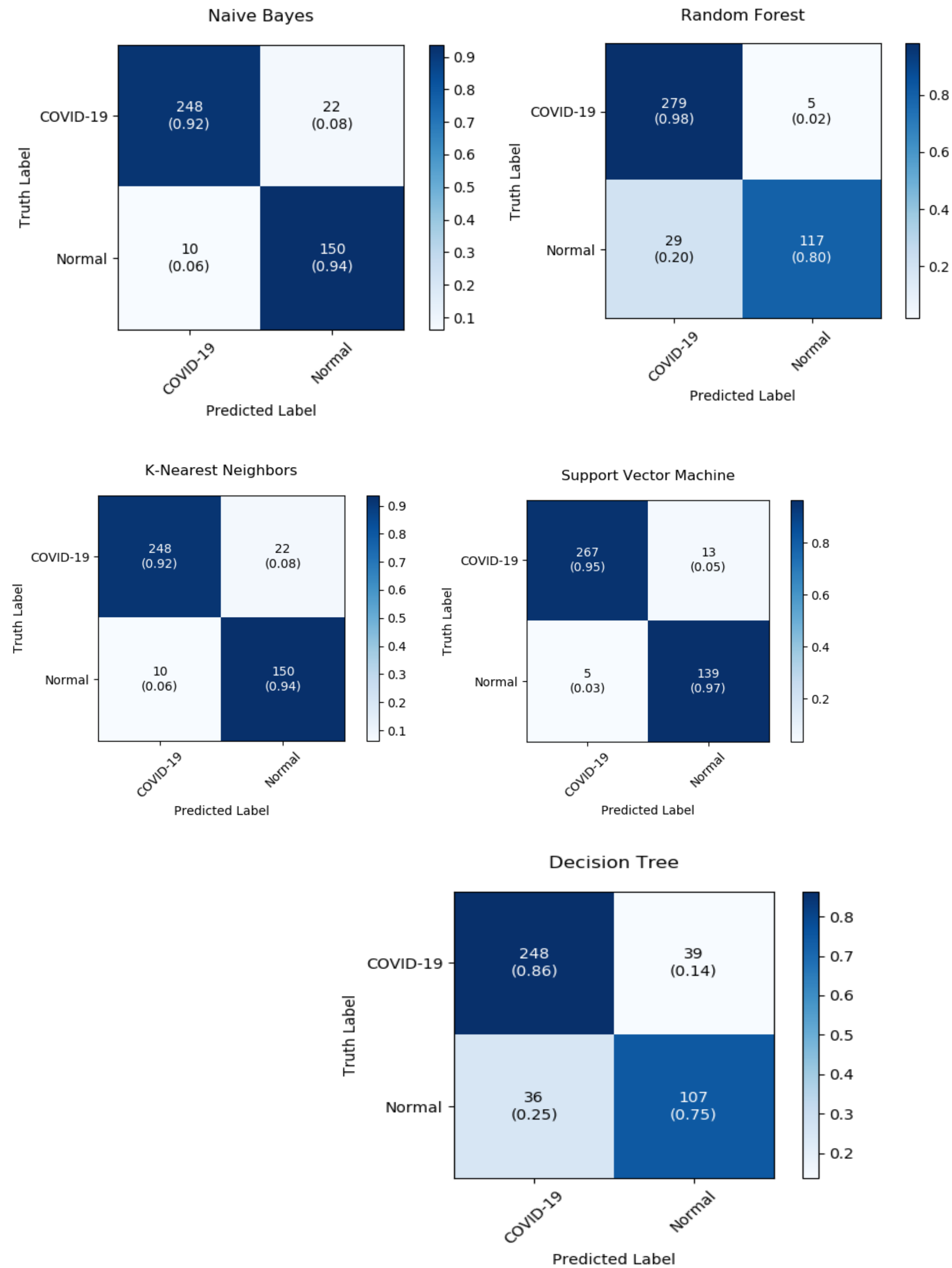



Fig. 5. Confusion Matrix of Different Machine Learning Technique

The performance of five machine learning techniques is illustrated in Figure 6 and Figure 7. The highest accuracy is obtained by a support vector machine of 96%. The K-Nearest Neighbor and Random Forest perform with a similar result with good accuracy of 92%, which is illustrated in Table 1 and Figure 6.

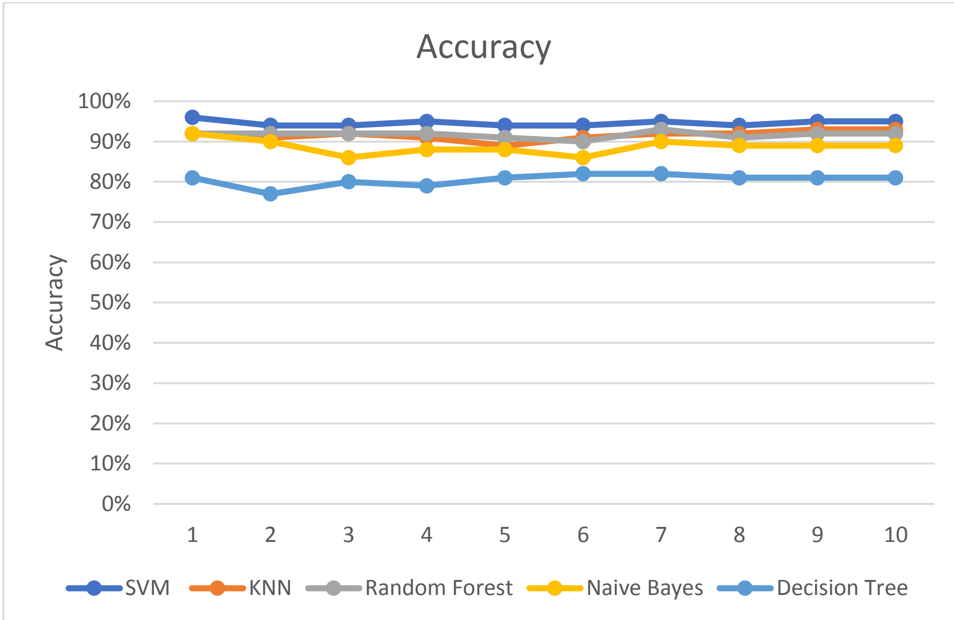
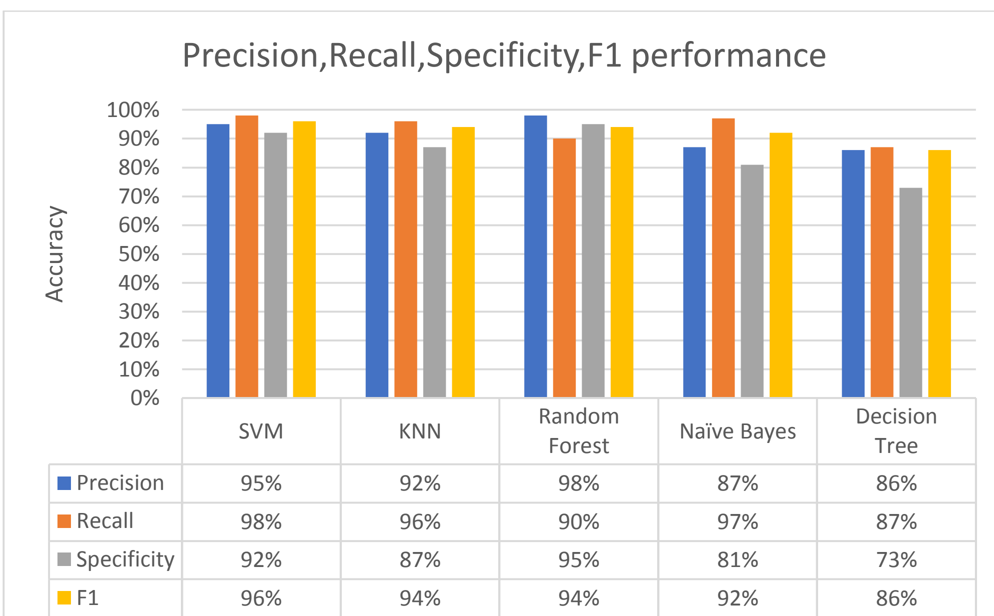



Fig. 6. Accuracy of five classification technique

The best performance model from the above machine learning technique is support vector machine, which is achieved the precision is 95%, recall is 98%, a specificity of 92% and the F1-score is 96%, while the lowest accuracy performance precision is 86%, recall is 87%, specificity is 73% and F1-score 86% which is illustrated in Figure 7.




| | SVM | KNN | Random Forest | Naïve Bayes | Decision Tree |
|---|---|---|---|---|---|
| Precision | 95% | 92% | 98% | 87% | 86% |
| Recall | 98% | 96% | 90% | 97% | 87% |
| Specificity | 92% | 87% | 95% | 81% | 73% |
| F1 | 96% | 94% | 94% | 92% | 86% |

Fig. 7. Performance of Machine Learning Technique (Precision, Recall, Specificity and F1-Score)

## 5. Conclusion

The coronavirus disease is widely spreading worldwide, which causes Severe Acute Respiratory Syndrome Coronavirus2 (SARS-CoV-2). Early prediction of COVID-19 can prevent the disease among other people. In this work, we utilized the benefits of using different machine learning techniques using chest X-ray images to distinguish automatically between the COVID-19 infected patients and normal Chest X-ray images. The performance of the different techniques shows that Support Vector Machine achieved the highest accuracy of 96% among the other machine learning techniques (K-Nearest Neighbor and Random Forest of 92%, Naïve Bayes of 90%, and Decision Tree of 82%). In the future, our intention is to increase the dataset to check the performance of the models and also used more classification techniques and feature extraction techniques to improve the performance accuracy, which helps the doctor to take a decision in clinical practice with high performance.

## References


[1] Yadav, M., Perumal, M., & Srinivas, M. (2020). Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*, *139*, 110050.

[2] Hassanien, A. E., Mahdy, L. N., Ezzat, K. A., Elmousalami, H. H., & Ella, H. A. (2020). Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine. *medRxiv*.

[3] Wu, J., Zhang, P., Zhang, L., Meng, W., Li, J., Tong, C., ... & Zhao, M. (2020). Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *medRxiv*.

[4] WHO Coronavirus Disease (COVID-19) Dashboard. (2020). Retrieved 2 October 2020, from https://covid19.who.int/?gclid=EAIaIQobChMI6eXWm6uW7AIVaRoGAB3otAEnEAAYASAAEgLmr_D_BwE

[5] Coronavirus. (2020). Retrieved 2 October 2020, from https://www.who.int/health-topics/coronavirus#tab=tab_1

[6] Ozkaya, U., Ozturk, S., & Barstugan, M. (2020). Coronavirus (COVID-19) Classification using Deep Features Fusion and Ranking Technique. *arXiv preprint arXiv:2004.03698*.

[7] Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *arXiv preprint arXiv:2003.13815*.

[8] Al-Karawi, D., Al-Zaidi, S., Polus, N., & Jassim, S. (2020). Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients. *medRxiv*.

[9] Alom, M. Z., Rahman, M. M., Nasrin, M. S., Taha, T. M., & Asari, V. K. (2020). COVID_MTNet: COVID-19 Detection with Multi-Task Deep Learning Approaches. *arXiv preprint arXiv:2004.03747*.

[10] Alqudah, A. M., Qazan, S., Alquran, H., Qasmieh, I. A., & Alqudah, A. COVID-2019 DETECTION USING X-RAY IMAGES AND ARTIFICIAL INTELLIGENCE HYBRID SYSTEMS.

[11] Sun, L., Liu, G., Song, F., Shi, N., Liu, F., Li, S., ... & Sun, L. (2020). Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *Journal of Clinical Virology*, 104431.

[12] Sethy, P. K., & Behera, S. K. (2020). Detection of coronavirus disease (covid-19) based on deep features. *Preprints*, *2020030300*, 2020.

[13] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *arXiv preprint arXiv:2004.09363*.

[14] Tartaglione, E., Barbano, C. A., Berzovini, C., Calandri, M., & Grangetto, M. (2020). Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data. *arXiv preprint arXiv:2004.05405*.

[15] Karim, M., Döhmen, T., Rebholz-Schuhmann, D., Decker, S., Cochez, M., & Beyan, O. (2020). Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images. *arXiv preprint arXiv:2004.04582*.

[16] Farid, A. A., Selim, G. I., Awad, H., & Khater, A. (2020). A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19). *Int. J. Sci. Eng. Res*, *11*(3), 1-9.

[17] Find Open Datasets and Machine Learning Projects | Kaggle. (2020). Retrieved 2 October 2020, from https://www.kaggle.com/datasets

[18] (2020). Retrieved 2 October 2020, from https://www.cs.auckland.ac.nz/courses/compsci373s1c/PatricesLectures/Image%20Filtering_2up.pdf

[19] Histogram equalization - Wikipedia. (2020). Retrieved 2 October 2020, from https://en.wiki-pedia.org/wiki/Histogram_equalization

[20] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

[21] Started, G., Tensorflow, K., Guide, R., Courses, O., (Old), C., & Consulting, A. (2020). Histogram of Oriented Gradients. Retrieved 2 October 2020, from https://www.learnopencv.com/histogram-of-oriented-gradients/

[22] Barstugan, M., Ozkaya, U., & Ozturk, S. (2020). Coronavirus (covid-19) classification using ct images by machine learning methods. *arXiv preprint arXiv:2003.09424*.

[23] Saeed, S., Shaikh, A., Memon, M. A., & Naqvi, S. M. (2018). Technique for Tumor Detection Upon Brain MRI Image by Utilizing Support Vector Machine. *Quaid-E-Awam University Research Journal of Engineering, Science & Technology, Nawabshah.*, *16*(01), 36-40.

[24] KNN using scikit-learn. (2020). Retrieved 2 October 2020, from https://towardsdatascience.com/knn-using-scikit-learn-c6bed765be75

[25] 1.6. Nearest Neighbors — scikit-learn 0.23.2 documentation. (2020). Retrieved 2 October 2020, from https://scikit-learn.org/stable/modules/neighbors.html

[26] KNN Classification using Scikit-learn. (2020). Retrieved 2 October 2020, from https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

[27] Naive Bayes classifier. (2020). Retrieved 2 October 2020, from https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[28] Naive Bayes Classifier. (2020). Retrieved 2 October 2020, from https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c
[29] Mandibules. (2020). Retrieved 2 October 2020, from https://en.wikipedia.org/wiki/Mandibules
[30] Understanding Random Forest. (2020). Retrieved 2 October 2020, from https://towardsdatascience.com/understanding-random-forest-58381e0602d2
[31] Decision Tree Classification in 9 Steps with Python. (2020). Retrieved 2 October 2020, from https://medium.com/pursuitnotes/decision-tree-classification-in-9-steps-with-python-600c85ef56de
[32] Decision Tree Classification in Python. (2020). Retrieved 2 October 2020, from https://www.datacamp.com/community/tutorials/decision-tree-classification-python
[33] Understanding Confusion Matrix. (2020). Retrieved 2 October 2020, from https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62