



Implementation of Machine Learning Techniques for the Classification of Lung X-Ray Images Used to Detect COVID-19 in Humans

Maad M. Mijwil

Computer Techniques Engineering Department, Baghdad College of Economic Sciences University,
Baghdad, Iraq

Received: 12/12/2020

Accepted: 16/3/2021

Abstract

COVID-19 (Coronavirus disease-2019), commonly called Coronavirus or CoV, is a dangerous disease caused by the SARS-CoV-2 virus. It is one of the most widespread zoonotic diseases around the world, which started from one of the wet markets in Wuhan city. Its symptoms are similar to those of the common flu, including cough, fever, muscle pain, shortness of breath, and fatigue. This article suggests implementing machine learning techniques (Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine) by Python to classify a series of chest X-ray images that include viral pneumonia, COVID-19, and healthy (Not infected) cases in humans. The study includes more than 1400 images that are collected from the Kaggle platform. The experimental outcomes of this study confirmed that the supported vector machine technique has high accuracy and excellent performance in the classification of the disease, as reflected by values of 91.8% accuracy, 91.7% sensitivity, 95.9% specificity, 91.8% F1-score, and 97.6% AUC.

Keywords: Machine Learning, X-Ray Images, COVID-19, Classification, Artificial Intelligence

تطبيق تقنيات التعلم الآلي لتصنيف صور الأشعة السينية للرئة لاكتشاف COVID-19 في جسم الإنسان

معد محسن مجو

¹قسم هندسة تقنيات الحاسوب، كلية بغداد للعلوم الاقتصادية الجامعية، بغداد، العراق

الخلاصة

يُعد COVID-19 (مرض فيروس كورونا -2019)، المعروف باسم CoV أو Coronavirus، مرضًا خطيرًا يسببه فيروس SARS-CoV-2. إنه أحد أكثر الأمراض الحيوانية المنتشرة حول العالم، بدءًا من أحد الأسواق الرطبة في مدينة ووهان. تتشابه أعراضه مع أعراض الأنفلونزا الشائعة، بما في ذلك السعال والحمى وألم العضلات وضيق التنفس والإرهاق. هذه الورقة تقترح بتطبيق تقنيات التعلم الآلي (Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine) بواسطة Python لتصنيف سلسلة من صور الأشعة السينية للصدر التي تحتوي على التهاب رئوي فيروسي، COVID-19.

*Email: maadalnaimiy@gmail.com

وسلامة (ليس مصاباً) للرئتين عند البشر. تتضمن هذه الدراسة أكثر من 1400 صورة تم جمعها من منصة Kaggle. أكدت النتائج التجريبية لهذه الدراسة أن تقنية آلة الناقل الداعمة لها دقة عالية وأداء ممتاز في تصنیف المرض، وهي 91.8% دقة، 91.7% حساسية، 95.9% خصوصية، 91.8% درجة F1 و 97.6% AUC.

1. Introduction

In 1918, a fatal pandemic called the Spanish flu (mother of all pandemics), swept through a third of the planet's population and killed between 40 to 50 million people within two years, according to the statistics of the US centre for disease control and prevention (CDC) and the world health organization (WHO) in Switzerland [1]. Today, after 100 years, another virus appears to be more powerful, more fatal, and faster to spread, threatening the lives of millions of people on this planet, which is named by the WHO as COVID-19[2] [3]. In March 2020, it was considered as a pandemic global communicable disease. Figure-1 shows people wearing masks in 1920 and 2020 (the difference is only a hundred years of time). The number of cases of COVID-19 disease is 71M (Recovered: 45.3M and Deaths: 1.59M), according to the WHO statistics until the end of this work (December 2020). In Iraq, the number of infections is increasing dramatically and noticeably among citizens of all ages and races. The control of this epidemic has become somewhat difficult to the point that the number of infections, since the spread of the virus in Iraq in February 2020 to the mid-December 2020, has reached 571K (Recovered: 502K and Deaths: 12,526). All this information and statistics are extracted from the official website of the ministry of health of Iraq as well as the WHO [4][5].

Chest radiological imaging techniques, such as X-radiation (X-ray) and computed tomography (CT), are favoured to diagnose COVID-19 in the early stages. Through these techniques, the percentage of infection in each person is distinguished [6]. The rapid spread of the disease in question and the increase in mortality rates in many nations reveal that an efficient treatment method should be generated. For this reason, measures to control the disease became obligatory, including early quarantine, diagnosis, and daily follow-up. At this point, artificial intelligence (AI) techniques can contribute to the above perspectives [7]. With a severe lack of specialists, while large similarities are observable between COVID-19 and traditional pneumonia, the AI powered auto-sensing model could be an important milestone towards dramatically reducing test time [8]. The solution system gives both a lower-cost and a more accurate diagnosis-treatment for COVID-19 as well as similar diseases like viral pneumonia. Figure- 2 (b) shows a high-resolution image of COVID-19 virus.



Figure -1 People wearing masks in 1920 (black and white photos) and 2020 (colour photo). These images are downloaded from Google and are freely available.

Many studies have been conducted in the literature using machine learning techniques (MLTs) for the purpose of diagnosis of patients with COVID-19 virus. Actually, many manuscripts and papers are pre-printed and can be found on the ResearchGate website. The ResearchGate has created a page only for COVID-19 research community; this page contains a large number of research papers and manuscripts, which are under continuous development and updating. For example, in Kumar *et al.* [9], a collection of chest x-rays is analysed using deep learning techniques (ResNet152) as well as machine learning for a group of people who suffer from the virus and another group of healthy people. This study achieved accuracy values of 97.3% by random forest technology and 97.7% by XGBoost technology to predict the disease. In a study conducted by Khanday *et al.* [10], a set of clinical reports were relied on to classify people with COVID-19 disease using MLTs (multinomial Naïve Bayes and

Logistic Regression) with the application of feature engineering parameters, such as frequency/inverse document frequency, Bag of words, and report length. In that study, the researchers were able to achieve high results in detecting the disease with greater than 96% accuracy. In a study conducted by Ahammed *et al.* [11], deep learning and ML approaches were applied to a set of chest x-ray images for patients with COVID-19. The database of this study includes images from the Kaggle and GitHub platforms. The results of this investigation described >94% accuracy, >95% AUC, >94% f-measure, >94% sensitivity, and >97% specificity. In a study conducted by Wang *et al.* from China, a technique of deep learning, namely Xception, was mixed with a technique of machine learning, which is the Support Vector Machine. They applied this idea to more than 1100 X-ray images of people with COVID-19 disease. Their study showed excellent results as it reached an accuracy of more than 99%, which is outstanding, while when applying the Xception technology alone, they achieved an accuracy of more than 96%. This paper is recommended to be viewed because it contains useful information [12].

There are also many studies in implementing deep learning techniques in the classifying of chest x-ray images of the disease caused by COVID-19 virus. For example, in a study conducted by Dansana *et al.* [13], the convolution neural networks method is applied with three classifiers, namely InceptionV2, VGG-19, and decision tree, to classify a collection of CT scan and X-ray images. The results of this study showed 91% of VGG-19, 78% of InceptionV2, and 60% of the decision tree model. In a study by Azemin *et al.* [14], the convolutional neural network architecture is applied with ResNet-101 classifier to detect COVID-19 in chest X-ray images. The outcomes of this study achieved an accuracy of higher than 71%, specificity of higher than 71%, and sensitivity of higher than 77%. Another study was conducted by Sekeroglu and Ozsahin from Turkey [15] who employed machine learning as well as deep learning techniques in detecting chest x-ray images for patients. This study includes 6100 images (Healthy:1583, Pneumonia:4292, and COVID-19:225). The results are as follows: Accuracy of >98%, specificity of >99% and sensitivity of >93%. In a recent study published in 2021, Castiglioni *et al.* proposed the application of deep learning in the analysis of chest X-rays of patients from Lombardy, Italy, where the number of images used reached 500 (COVID-19 and non-COVID-19 images). This study reached results of classification as follows: 0.78% sensitivity, 0.82% specificity, and 0.89% AUC [16]. There are still many new articles and papers written and published every day in the hope of serving to get effective treatment to overcome this epidemic.

The main contribution of this scenario is to classify a set of chest X-ray images taken from Kaggle.com, by applying very beneficial classic machine learning techniques in predicting the COVID-19 virus, and analyse the results of these techniques by accuracy, sensitivity, specificity, F1-score, and AUC. These technologies will play an essential and influential role in identifying, predicting, and preventing illnesses caused by coronavirus. Fortunately, machine learning, as one of the well-known applications of AI, has been extensively applied to many COVID-19 datasets in many articles.

The rest of the article is organised as follows: Section 2 covers a description of the COVID-19 and how it began and spread throughout the world, its symptoms, and the structure of this pandemic. In Section 3, materials and methods are discussed. In Section 4, experimental outcomes of ML techniques are presented. The last section discusses the conclusions and future work for improvements.

2. COVID-19 Pandemic Structure

The novel coronavirus (COVID-19) epidemic is a hard episode that began in Huanan Seafood Market in Jianghan District, Wuhan, China, in December 2019. It was described to be transmitted from bats to individuals, being spread around the world. It became a fatal epidemic which health systems in many nations failed to control. The most prominent symptoms and signs of infection are dry coughs, fever, sore throat, headache, muscle pain, weakness, diarrhoea, and shortness of breath [17]. In more advanced cases, it causes severe pneumonia, causing inflammation in the lungs due to oxygen difference and multiple organ failure. Significantly, this disease has more dangerous and corrosive effects for those with chronic illnesses, those with weak resistance or immune system, smokers, and the elderly [18].

All over the world, the number of patients with COVID-19 virus is developing day by day. Even well-developed countries like Italy, the United States, France, Spain, and the United Kingdom could not shield themselves sufficiently and are hugely influenced by this pandemic. In the light of all this

information, early diagnosis of the disease is of great interest to stop the pandemic effectively or at least to reduce the possible damage of the virus, taking into account the health system in each country. The incubation period lasts from one to fourteen days, most often around four or five days. Symptoms appear gradually: muscle pain, headache, and fatigue [19]. Dry cough and fever then occur at the height of the disease, with possible chest pain and problems in breathing [20]. In some victims, these symptoms are accompanied by a sore throat, while the psychological state of victims also deteriorates. Some victims do not present any symptoms or signs. This category of people is considered the most dangerous in transmitting the virus to their families and friends or to anyone in the community. Other clinical symptoms/ signs have been reported as the sudden loss of smell and taste, diarrhoea, or the appearance of frostbite on the extremities of the hands and feet [21].

Coronaviruses have rounded morphology and diameter of 100-150nm (about 600 times smaller than the diameter of a human hair!) [22]. Starting from the outermost layer and progressing gradually towards the inside of the virus, it is possible to notice different structures as shown in Figure- 2 (a) and listed below:

- **S(Spike)-Glycoprotein [23]**: the virus shows projections on its surface, with a length of about 20 nm. Three S-Glycoproteins are joined together to make up a trimer; the trimers of this protein form the structures which, taken together, resemble a corona surrounding the virion.
- **M(Membrane)-Protein [24]**: this protein actively interacts the virus with the ribonucleic acid (RNA) protein complex.
- **Hemagglutinin Esterase (HE) [25]**: this coating protein, smaller than S-Glycoprotein, plays an important role during the virus release phase within the host cell.
- **E-Protein [26]**: this protein helps the S-Glycoprotein to attach itself to the membrane of the target cell.
- **Envelope [27]**: it is the envelope of the virus, consisting of a membrane that the virus "inherits" from the host cell after having infected it.
- **RNA and N(Nucleocapsid)-Protein [28]**: the Coronavirus genome is made up of a single strand of large size positive polarity RNA (from 27 to 32kb in the different viruses). RNA gives rise to seven viral proteins which are associated with the N-protein, which enhances its durability.

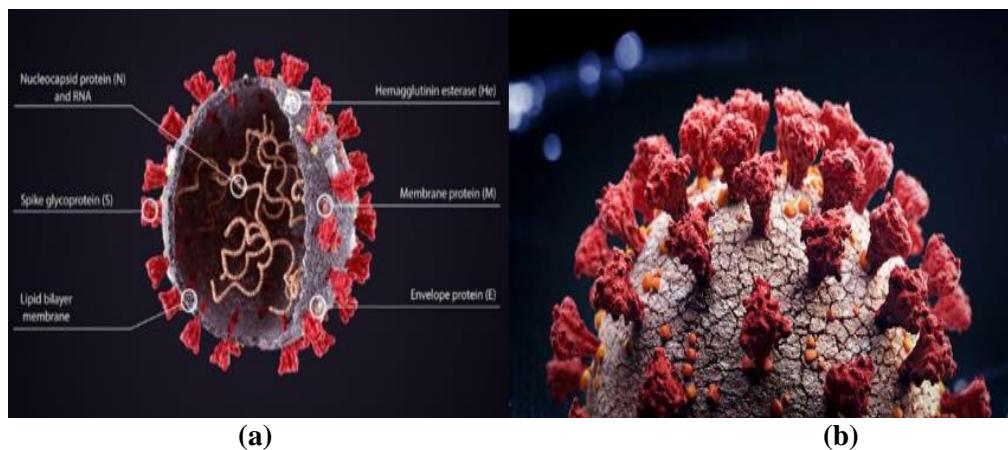


Figure -2 (a) Structure of the COVID-19, (b) High-resolution image of the virus from Goggle images (these images are open to all users).

3. Materials and Methods

3.1 COVID-19 Dataset

In this section, machine learning techniques are introduced, but before that, we must know the difference between a lung infected with a Coronavirus (COVID-19), traditional pneumonia, and normal (healthy). This is illustrated by Figure- 3 that demonstrates chest X-ray images in victims. Figure- 3 (a) exhibits clear lungs, which are normal with no unusual zones (not infected). Figure- 3 (b) presents viral pneumonia with a more scattered pattern in both lungs. Figure- 3 (c) shows lungs which are infected with the Coronavirus (COVID-19). All of these images are collected from Mooney [29] and Khoong [30], while some images were used from Google (to enhance the code), with the file size reached more than 2 GB.

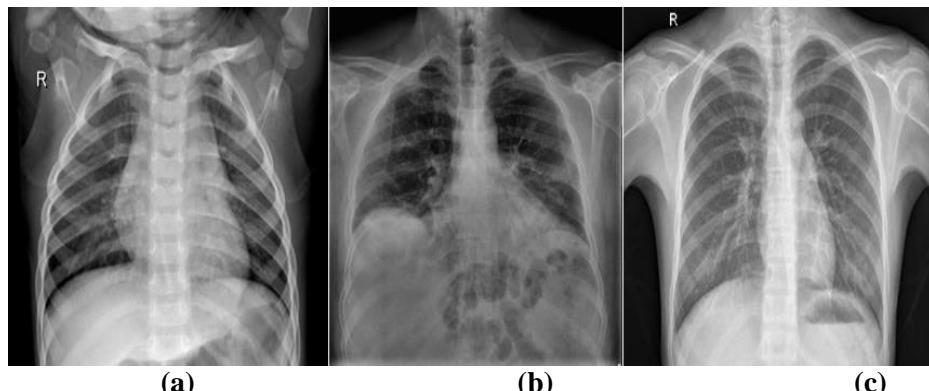


Figure -3 Example of X-Ray images for lungs in humans. (a) Normal lungs, (b) Viral pneumonia, and (c) COVID-19.

Table-1 shows the total dataset (train sets & test set) collected from this platform. On this platform, a database called (CORD-19) has been created and includes more than 29k chapters and more than 13k full-text papers about coronaviruses (SARS-CoV-2, COVID-19). This openly accessible dataset is presented to the global study community to implement recent advances in natural language processing and other artificial intelligence techniques to generate new opinions in support of the ongoing fight against this dangerous disease. In addition to applying the machine learning techniques in classifying and predicting the infection of COVID-19 along with viral pneumonia cases, these techniques have also been applied in predicting the survival rates of coronavirus patients.

In this study, 70% of the data in the dataset is applied to train the model and 30% to test the model. These images are of different sizes and with high-resolution, where the dimensions are 2024 * 1826 in JPEG format. For preparing the final experiments, the format of all the images is converted to Portable Network Graphics (PNG) while preserving the characteristics of the images. Furthermore, each of the viral pneumonia, COVID-19, and healthy class images are resized to 224*224. The working mechanism depends on analysing the test images for all the classes used to classify them, with finding each classifier's accuracy in determining the disease in the patient's lung and giving the best accuracy result.

Now, the techniques that are applied in this study to classify X-ray images to detect COVID-19 disease will be discussed in brief, and they are illustrated below. Figure- 4 illustrates the steps of work, which are the input stage, the models selection (feature extraction, feature selection, and classifier), and the output stage.

Table 1 X-ray images dataset utilised in this work is a training set and a test set.

Dataset Bifurcation	Classes		
	COVID-19	Pneumonia	Normal
Training Set	389	401	397
Test Set	113	113	113
Training+ Test set	420	514	510
Final total: 1445			

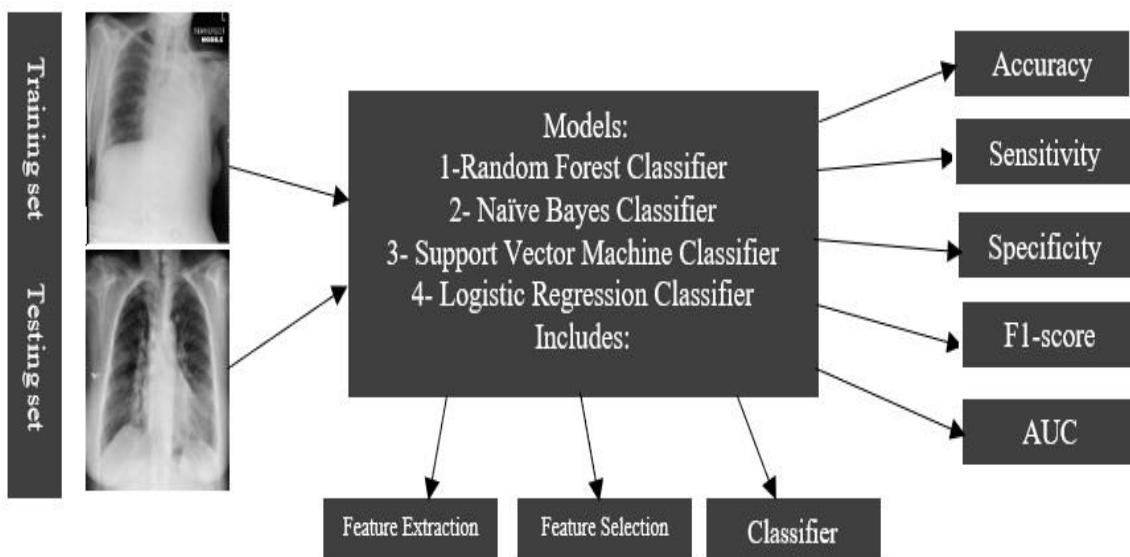


Figure 4-The stages of work that start from entering information (images) to presenting results for each technique.

3.2 Machine Learning Techniques

In this sub-section, the machine learning techniques applied in this work are briefly described, along with the working mechanism, and the features of each one.

- Random Forest (RF) [31]

It is a supervised classification model based on majority voting. Simply put, this classifier includes electing a random set of properties and generating a classifier with a boot sample of training data. In this way, a maximum number of decision trees is created and weighted voting is applied to eventually assign an unknown value to a class. This technique is characterized by treating the absent values through two methods; the first is by using the median values of the data to replace the continuous variables and the second is by computing the proximity-weighted average of absent values. Forest technique continues to run to get the job done, does not suffer from any overfitting, and never stops, making it very effective and trustworthy. The reason for that is that it takes the average of all predictions, which eliminates biases.

- Naïve Bayes (NB) [32]

This theory goes back to the reverend Thomas Bayes (English statistician), the person who studied probability and binomial distributions in the eighteenth century. NB is a pure probabilistic classifier based on the bayes theorem. Simply put, this classifier assumes that the predictor variables are independent of each other. In other words, the presence of a certain feature in a dataset is not at all related to the presence of any other feature. In addition, this classifier works well in many complex real-world situations. The power of this technique is that it assumes that the features of the data point are completely independent of each other. This technique relies on its work on probabilities to achieve certain goals to make predictions and classifications of data. This factor makes this technique unique in creating an intelligent system of extracting correct, accurate, and reliable data for the user. NB ignores all unrelated features and considers them strange. Another feature that some think is poor, but which is preferred by others, is that it works with huge datasets.

- Support Vector Machine (SVM) [33]

It is a supervised classification model that runs with powerful algorithms to recognize patterns of data behaviour. It is a widely executed tool in support of the medical image classification process. SVM is based on some nice simple ideas and provides a clear understanding of what the things learned from the examples mean. It can lead to high-performances in practical applications. Vladimir N. Vapnik launched this technique in 1995. SVM is based on analytical equations with different kernel function types to display separable nonlinear samples in high-dimensional domains. Kernel functions play a significant role in this technique of converting data from linear to nonlinear. In recent years, this technique has been widely utilised in the medical field, especially in applications including predicting brain diseases, Alzheimer's disease, hepatitis, and now analysing lung images (X-ray & CT-scan

images) of patients with COVID-19 disease. This technique operates on areas of high-dimensions and begins to work when the margins are clear and separable. One of its drawbacks is that it is not suitable for a large data set, which is one of the reasons that led to the success of this technique in this work.

- Logistic Regression (LR) [34]

It is a powerful and well-established method in the field of statistics and biomedicine. LR compares categorical results and an explanatory variable. It is a prediction model that can be used when the target variable is a categorical variable with two categories; for example active or inactive, healthy or unhealthy, winning or losing. The speed in implementation is the most valuable feature of the work of this technique compared to other techniques that are supervised, such as SVM. The methods applied in this technique with complex relationships between variables are very simplified. Its performance tends to be impaired if the decision boundaries are non-linear. LR's mechanism depends on the log odds ratio instead of the normal probabilities and the iterative maximum likelihood method instead of the least-square to fit the final model of classification and prediction. This means that the researcher has a great freedom in using this technique, especially when the classes are different and contain unequal arrays.

4. Experiments Results

This section shows an analysis of the performance of machine learning techniques based on the results obtained by Python v3.8.0 and Spyder IDE v.4.2.1 with CPU: 2.30GHz Core i5-8300H, Graphics Cards: NVIDIA GeForce GTX 1050 with 4GB graphics (gaming), RAM: 8GB, and running on Windows 10 Home (OS Build 19041.488). In addition, a comparison is made between these techniques to reach the best technique in analysing these images, know their accuracy in recognising the type of disease in the patient's lung, and determine the technique whose accuracy is not appropriate in analysing these images. Tables 2, 3, 4, 5, 6 and Figure-5 present the results achieved. The performance of machine learning techniques is estimated using the classical metrics of accuracy (*Acc.*), sensitivity (*Sen.*), specificity (*Spe.*), and F1-score, as shown in the mathematical formulas (1), (2), (3), and (4), respectively.

$$Acc. = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$Sen. = \frac{TP}{TP+FN} \quad (2)$$

$$Spe. = \frac{TN}{TN+FP} \quad (3)$$

$$F1 - score = \frac{2*TP}{2*TP+FP+FN} \quad (4)$$

where TN refers to the true negative, TP refers to the true positive, and FN and FP denote the false negative and the false positive, respectively. AUC is a value that uses the parameters of false-positive rate (FPR) and true-positive rate (TPR). It determines the performance of a classification model at all classification thresholds. Area under the curve for ROC (AUC-ROC) is an efficient measure to check the effectiveness of ML classifiers and it can be measured using formulas 5, 6, and 7. If the AUC is higher, it means that the classifier has a higher distinguishing ability, which can strongly distinguish the classes of images used and determine the presence or absence of viruses and the types of viruses in human lungs. Table 7 exhibits the time consumed by each model in classifying X-ray images. This table covers the training time used to analyse test images and the time used to analyse one image sample of each class.

$$FPR = \frac{FP}{FN + FP} \quad (5)$$

$$TPR = \frac{TP}{FN + TP} \quad (6)$$

$$AUC = \int_{x=0}^1 TPR(FPR^{-1}(x))dx \quad (7)$$

Table 2-Results of the ML models for classification among Normal, COVID-19, and Viral Pneumonia images

Models	Performance Metrics				
	Accuracy	Sensitivity	Specificity	F1-score	AUC
NB	89.4%	89.4%	94.7%	89.4%	97.3%
RF	90.3%	90.3%	95.3%	90.2%	97.5%
LR	91.2%	91.1%	95.8%	91.2%	97.1%
SVM	91.8%	91.7%	95.9%	91.8%	97.6%

Table 3 Confusion matrix for the performance of NB

Predicated Class				
Actual Class		COVID-19	Normal	Pneumonia
	COVID-19	96	2	15
Normal		2	1.1e+02	0
Pneumonia		17	0	96

Table 4 Confusion matrix for the performance of RF

Predicated Class				
Actual Class		COVID-19	Normal	Pneumonia
	COVID-19	95	2	16
Normal		1	1.1e+02	0
Pneumonia		14	0	98

Table 5 Confusion matrix for the performance of LR

Predicated Class				
Actual Class		COVID-19	Normal	Pneumonia
	COVID-19	92	1	20
Normal		2	1.1e+02	0
Pneumonia		7	0	1.1e+02

Table 6 Confusion matrix for the performance of SVM

Predicated Class				
Actual Class		COVID-19	Normal	Pneumonia
	COVID-19	96	2	17
Normal		2	1.1e+02	0
Pneumonia		9	0	1e+02

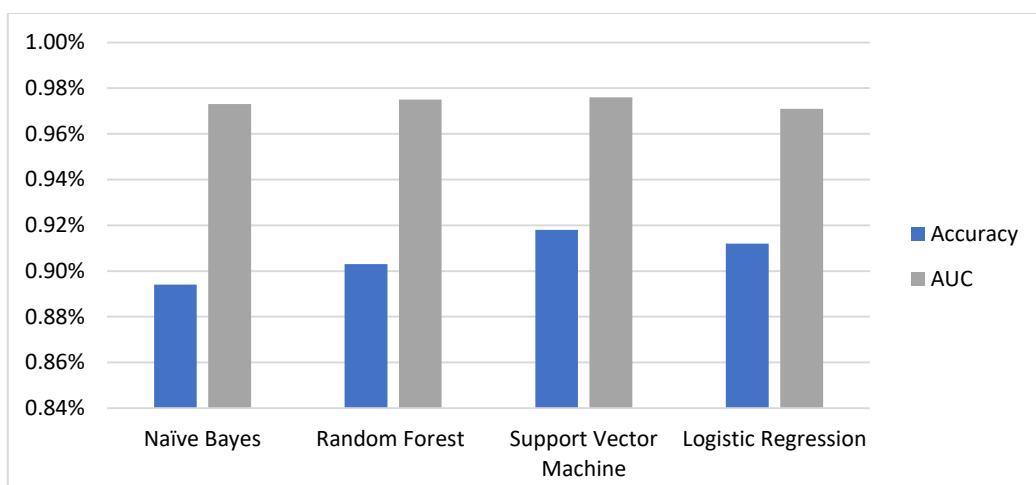
**Figure 5**-The behaviours of accuracy and AUC of each classifier in the classification manner of images

Table 7-Training time and testing time for the models

Models	Training Time (sec)	Testing Time for one sample (sec)
NB	2639.316755	0.01322173095
RF	3606.886011	0.022614044305
LR	3348.548371	0.02040854124
SVM	19708.88222	0.1363421233

5. Conclusions and Future Work

In this article, machine learning techniques are employed to classify a set of chest X-ray images of people with pneumonia and COVID-19, as well as images of healthy people. After several investigations, it has been found that the support vector machine technique is the most beneficial in terms of accuracy in detecting COVID-19, while the logistic regression technique has good and acceptable results in terms of accuracy. The performance of the random forest technique is quite good. The worst performance is that of the naïve bayes technique that does not give convincing and satisfactory results in this article. This article is concentrated on the values of accuracy and AUC because these values are essential in measuring the level of implementation of the Machine Learning classifiers among the three classes of COVID-19, normal, and pneumonia cases. In the future, deep learning techniques will be applied to these X-ray images of the injured human lung, and an analysis of the effects of these techniques will be made. Hopefully, developments in these techniques will finally contribute to the overall global efforts to overcome this crisis.

6. Disclosure and conflict of interest

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Reference

1. Y. A. Helmy, M. Fawzy, A. Elaswad, A. Sobieh, S. P. Kenney, and A. A. Shehata, **2020**. "The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control," *Journal of Clinical Medicine*, **9**, ID. 1225, pp. 1-29, April 2020. <https://doi.org/10.3390/jcm9041225>
2. Mijwil M. M. and Al-Zubaidi E. A., **2021**. "Medical Image Classification for Coronavirus Disease (COVID-19) Using Convolutional Neural Networks," *Iraqi Journal of Science*, **62**(8), In Press, August 2021.
3. Ministry of Health of Iraq, <https://www.facebook.com/MOH.GOV.IQ/>
4. World Health Organization website, "WHO Coronavirus (COVID-19) Dashboard-2021", <https://covid19.who.int/table>
5. Z. Sun, N. Zhang, Y. Li, and X. Xu, **2020**. "A systematic review of chest imaging findings in COVID-19," *Quantitative Imaging in Medicine and Surgery*, **10**(5): 1058–1079, May 2020. <https://doi.org/10.21037/qims-20-564>
6. R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, **2020**. Artificial Intelligence (AI) applications for COVID-19 pandemic, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**: 337-339, April 2020. <https://doi.org/10.1016/j.dsx.2020.04.012>
7. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, **2020**. "Automated Detection of COVID-19 Cases Using Deep Neural Networks with X-Ray Images," *Computers in Biology and Medicine*, **121**, ID. 103792, pp. 1-9, April 2020. <https://doi.org/10.1016/j.combiomed.2020.103792>
8. Kumar, R. Arora, V. Bansal, V. J. Sahayashela, H. Buckchash, J. Imran, N. Narayanan, G. N. Pandian, and B. Raman, **2020**. "Accurate Prediction of COVID-19 Using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers," *MedRxiv Preprint*, **1**: 1-10, April 2020. <https://doi.org/10.1101/2020.04.13.20063461>
9. A. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. M. Din, **2020**. "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal of Information Technology*, pp.1-9, Jun 2020. <https://doi.org/10.1007/s41870-020-00495-9>

10. K. Ahammed, S. Satu, M. Z. Abedin, A. Rahaman, and S. M. S. Islam, "Early Detection of Coronavirus Cases Using Chest X-ray Images Employing Machine Learning and Deep Learning Approaches," This manuscript is preprint, pp.1-15, June 2020.
11. Wang D., Mo J., Zhou G., Xu L., and Yajun Liu, **2020**. "An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images," *Plos one*, **15**(11):1-15, November 2020. <https://doi.org/10.1371/journal.pone.0242535>
12. D. Dansana, R. Kumar, A. Bhattacharjee, D. J. Hemanth, D. Gupta, A. Khanna, O. Castillo, **2020**. "Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm," *Soft Computing-Springer*, pp.1-9, Auguste 2020. <https://doi.org/10.1007/s00500-020-05275-y>
13. M. Z. C. Azemin, H. Hassan, M. I. M. Tamrin, and M. A. M. Ali, **2020**. "COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings, *International Journal of Biomedical Imaging*, vol. 2020, ID. 8828855, pp. 1-7, Auguste 2020. <https://doi.org/10.1155/2020/8828855>.
14. B. Sekeroglu, and I. Ozsahin, **2020**. "Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks," *SLAS Technology*, pp. 1-13, September 2020. <https://doi.org/10.1177/2472630320958376>
15. I. Castiglioni, D. Ippolito, M. Interlenghi, C. B. Monti, C. Salvatore, S. Schiaffino, A. Polidori, D. Gandola, C. Messa, and F. Sardanelli, **2020**. "Machine learning applied on chest x-ray can aid in the diagnosis of COVID-19: a first experience from Lombardy, Italy," *European Radiology Experimental*, **5**(7):1-10, February 2021. <https://doi.org/10.1186/s41747-020-00203-z>.
16. A. Hafeez, S. Ahmad, S. A. Siddqui, M. Ahmad, and S. Mishra, **2020**. "A Review of COVID-19 (Coronavirus Disease-2019) Diagnosis, Treatments and Prevention," *Eurasian Journal of Medicine and Oncology*, **4**(2): 116–125, April 2020, <https://doi.org/10.14744/ejmo.2020.90853>
17. F. A. Rabi, M. S. Al Zoubi, G. A. Kasasbeh, D. M. Salameh, and A. D. Al-Nasser, **2020**. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far," *Pathogens*, **9**(231): 1-14, March 2020. <https://doi.org/10.3390/pathogens9030231>
18. J. Dai, L. Yang, and J. Zhao, **2020**. "Probable Longer Incubation Period for Elderly COVID-19 Cases: Analysis of 180 Contact Tracing Data in Hubei Province, China," *Risk Management and Healthcare Policy*, **13**: 1111–1117, August 2020. <https://doi.org/10.2147/RMHP.S257907>
19. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich N, and J. Lessler, **2020**. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application," *Annals of Internal Medicine*, **172**(9): 577-582, May 2020. <https://doi.org/10.7326/M20-0504>
20. R. W. Miller, **2020**. there's a new symptom of coronavirus, doctors say: Sudden loss of smell or taste, USA Today, May 2020, Article available on: <https://www.usatoday.com/story/news/health/2020/03/24/coronavirus-symptoms-loss-smell-taste/2897385001/>
21. T. B. Knudsen, T. N. Kledal, O. Andersen, J. Eugen-Olsen, and T. B. Kristiansen, **2003**. "Severe Acute Respiratory Syndrome – a New Coronavirus from the Chinese Dragon's Lair," *Scandinavian Journal of Immunology*, **58**(3): 277–284, September 2003. <https://doi.org/10.1046/j.1365-3083.2003.01302.x>.
22. A. Mittal, K. Manjunath, R. K. Ranjan, S. Kaushik, S. Kumar, and V. Verma, **2020**. "COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2," *Plos Pathogens*, **16**(8):1-19, August 2020. <https://doi.org/10.1371/journal.ppat.1008762>
23. S. Mukherjee, D. Bhattacharyya, and A. Bhunia, **2020**. "Host-membrane interacting interface of the SARS coronavirus envelope protein: Immense functional potential of C-terminal domain," *Biophysical Chemistry*, **266**(106452):1-14, November 2020. <https://doi.org/10.1016/j.bpc.2020.106452>.
24. C. H. Kim, **2020**. "SARS-CoV-2 Evolutionary Adaptation toward Host Entry and Recognition of Receptor O-Acetyl Sialylation in Virus–Host Interaction," *International Journal of Molecular Sciences*, **21**(12): 1-34, January 2020. <https://doi.org/10.3390/ijms2112454951>.
25. I. Astuti , and Ysrafil, **2020**. "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**: 407-412, April 2020. <https://doi.org/10.1016/j.dsx.2020.04.020>

26. L. Mousavizadeh , and S. Ghasemi. **2020.** “Genotype and phenotype of COVID-19: Their roles in pathogenesis,” *Journal of Microbiology, Immunology and Infection*, In Press, March 2020.
27. S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Y. Yan, C. Zhang, H. Shan, and S. Chen, **2020.** “Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites,” *Acta Pharmaceutica Sinica B*, **10**(7): 1228-1238, July 2020. <https://doi.org/10.1016/j.apsb.2020.04.009>
28. P. Mooney, “Chest X-Ray Images (Pneumonia),” Datasets: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
29. W. H. Khoong, “COVID-19 Xray Dataset (Train & Test Sets),” Datasets: <https://www.kaggle.com/khoongweihao/covid19-xray-dataset-train-test-sets>
30. R. Pourdarbani, S. Sabzi, M. H. Hernández, J. L. H. Hernández, G. G. Mateos, D. Kalantari, and J. M. M. Martínez, **2019.** “Comparison of Different Classifiers and the Majority Voting Rule for the Detection of Plum Fruits in Garden Conditions,” *Remote Sensing-MDPI*, **11**: 1-17, October 2019. <https://doi.org/10.3390/rs11212546>
31. Z. Zhang, **2016.** “Naïve Bayes classification in R,” *Big-data Clinical Trial Column*, **4**(12): 241-246, June 2016. <https://doi.org/10.21037/atm.2016.03.38>
32. J. Nalepa and M. Kawulok , **2018.** “Selecting Training Sets for Support Vector Machines: A Review,” *Artificial Intelligence Review-* Springer Nature, **52**: 857–900, January 2018. <https://doi.org/10.1007/s10462-017-9611-1>.
33. M. Kirişci, **2019.** “Comparison of Artificial Neural Network and Logistic Regression Model for Factors Affecting Birth Weight,” *SN Applied Sciences-Springer Nature*, **378**: 1-9, March 2019. <https://doi.org/10.1007/s42452-019-0391-x>