Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
Published Online First: May 2022 **2022, 19(6): 1356-1361** **E-ISSN: 2411-7986**

DOI: https://dx.doi.org/10.21123/bsj.2022.5965

Diagnosing COVID-19 Infection in Chest X-Ray Images Using Neural Network

***Salim M. Zaki*[1*]** ***Mustafa Musa Jaber*[2,3]** ***Mohammed A. Kashmoola*[4]**

[1] Department of Computer Science, Dijlah University College (DUC), Baghdad, Iraq
[2] Department of Medical Instruments Engineering Techniques, Al-Farahidi University, Baghdad, 10021, Iraq
[3]Al-Turath University College, Baghdad, Iraq
[4] Department of Computer Science, College of Education, University of Al-Hamdaniya, Nineveh, Iraq
*Corresponding author: salimzki@gmail.com
E-mail addresses: mustafa.musa@duc.edu.iq , mkashmola@uohamdaniya.edu.iq

Received 24/10/2020, Accepted 22/9/2021, Published Online First 20/5/2022, Published 1/12/2022

Abstract:

With its rapid spread, the coronavirus infection shocked the world and had a huge effect on billions of peoples' lives. The problem is to find a safe method to diagnose the infections with fewer casualties. It has been shown that X-Ray images are an important method for the identification, quantification, and monitoring of diseases. Deep learning algorithms can be utilized to help analyze potentially huge numbers of X-Ray examinations. This research conducted a retrospective multi-test analysis system to detect suspicious COVID-19 performance, and use of chest X-Ray features to assess the progress of the illness in each patient, resulting in a "corona score." where the results were satisfactory compared to the benchmarked techniques. This research results showed that rapidly evolved Artificial Intelligence (AI) -based image analysis can accomplish high accuracy in detecting coronavirus infection as well as quantification and illness burden monitoring.

Keywords: COVID-19, Deep Learning, Image Processing, Neural Network, X-Ray Image Processing

Introduction:

The spread of coronavirus, COVID-19 has astounded the world with its fast spread, expected harmfulness, with large effect on the lives of billions of individuals from both wellbeing and a financial point of view. As of this composition, there are roughly 93,158 affirmed instances of which 80,270 are in "Territory China" with 3,198 deaths, a death pace of 3.4%[1].

According to Jiang F et al[1] and World Health Organization (WHO)[2], breathing difficulty is one of the main indications of COVID-19, it could be diagnosed by chest X-ray images. Similarly, CT scans can also show the symptoms of COVID-19. Therefore, analyzing medical images of patients' chests could help in the detection of COVID-19[3] . Processing these images can substitute the lack of kits used by medical staff for diagnosing the diseases especially when many countries faced the limitation of the available number of kits compared to the high number of infected or suspicious to be infected people. The availability of x-ray and CT scan devise in most of the hospitals and medical laboratories make this technique of identifying COVID-19 with the assistance of machine learning an easy and low-cost technique of diagnosing particularly when the preliminary symptoms such as fever do not appear on patients[4].

Relevant studies from data analysis based on COVID-19 identification are of very limited nature. Where most researchers focus on medical images such as X-rays and CT scans. While other studies are working on data predictions for modeling the spread of COVID-19 infection with a layered perceptron. Another study uses data mining techniques for COVID-19 data. These different methods all work towards one target that is diagnosing COVID-19 infection[5,6].

Yicheng, et al[7] studied many cases of patients with COVID-19 infection and compared results of CT scan and RT-PCR (real-time polymerase chain reaction). Many signs of infection were discovered in images of lungs for people with fever for few days or coughing. The CT scan with image processing techniques proved the diagnoses of COVID-19 infection in early stages of infection. Yicheng, et al[7] proposed an affectability of non-differentiated chest CT for detection of COVID-19 disease of 98% contrasted with introductory RT-PCR affectability (consequences of the primary RT_RPR trial) of 71%. Cases featured in their paper showed either diffuse or central ground-glass opacities. This absence of affectability on starting RT-PCR testing was likewise depicted in another investigation by Xie *et al* [8] who showed that the low number of patients with negative RT-PCR values of the contagion regardless of chest CT discoveries average of viral pneumonia recommending the utilization of chest CT to diminish bogus negative lab contemplates. Bernheim *et al*[9] analyzed 121 chest CT concentrates from four centers in China that were acquired in the early, and late phases of contamination. They additionally portrayed ground glass opacities as normal for the infection, especially reciprocal and fringe ground-class and consolidative pneumonic opacities. They noted more prominent seriousness of ailment with expanding time from the beginning of indications and they depicted later indications of ailment to incorporate "more prominent lung association, straight opacities, "insane clearing" design, and the "opposite corona" sign. There was reciprocal lung contribution in 28% of early patients, 76% of patients with middle cases, and around 88 percent of patients in late steps (from 6 to 12 days) of sickness.

When a choice was taken to use X-Ray as these recent findings indicate for patient treatment or screening, a need for quick assessment of a huge number of imaging studies immediately increases. Artificial intelligence technology, particularly tools for image analysis of deep learning, can be built to help radiologists, quantification and tendency analysis of the data. AI applications can parallel examining several cases to determine whether a chest X-Ray image shows any lung abnormalities. Where the software indicates that a risk is significantly increased of disease, a radiologist or clinician may flag the case for potential treatment/quarantine for further examination. Upon checking and reviewing, these systems or variations may become important in diagnosing and controlling patients with COVID-19.

An artificial intelligence algorithm can take the task of diagnosing COVID-19 infections with processing images from medical devices. Developing such an algorithm should include the phases: I. Data-collection phase where data collected from an online source; II. Training the algorithm on data set. III. The testing section to test the output and compare it to a real classification in order to examine the achievement of the developed algorithm.

Another malady, for example, the coronavirus, datasets are quite recently being recognized and commented on. There are exceptionally constrained information sources just as restricted ability in naming the information explicit to this new strain of the infection in people. As needs are, it isn't certain that there are sufficient guides to accomplish clinically significant learning at this beginning phase of information assortment regardless of the inexorably basic significance of this product. We speculate that AI-based techniques can be quickly evolved utilizing the capacity to change and adjust existing AI models and join them with beginning clinical comprehension to address the new difficulties and new class of COVID-19.

The list of contributions presented in this research are:

- Utilizing artificial intelligence technique to classify x-ray images of patients' lungs infected with the COVID-19 virus. The classification of images make the infection clear of the virus in the lungs images.
- Building a deep learning algorithm to assist specialists in the hospital to identify COVID-19 positive cases from negative ones with less risk.
- High accuracy of identifying infected samples with COVID-19.

The proposed idea shows the gaps in the area of COVID-19 and the usage of AI techniques to motivate researchers in the field of artificial intelligence to keep developing ideas for making diagnoses of contagious diseases earlier and more accurately. Therefore, the objective of this research is creating artificial intelligence-based computerized X-Ray picture investigation apparatuses for identification, measurement, and chasing Coronavirus and show that they can separate patients diagnosed with coronavirus from individuals without the illness.

Methodology:

Neural Network:

Machin Learning (ML) is useful in making machines classifying things inside images smarter than other traditional techniques. These AI techniques make advances in different fields such as diagnosing medical sicknesses and in the industry[10-12]. The smart method of neural networks in learning and identifying makes it superior especially in generating semantic features from data entered to it[13,14]. Machine learning is used in analyzing data on social media platforms such as Twitter for management strategies development. Yet, the need to utilize ML for identifying COVID-19 patients is a global demand[15,16]. The high risk of COVID-19 spread and infection is due to easy transmission primarily through coughing which makes it a real threat to people around the world[17]. This motivated authors of this research to rely on a neural network and utilize it for detecting COVID-19 infection in x-ray images and classifying input images as infected or normal.

Data Set:

Data set is the fuel of neural network for deep learning and coming without realistic output. COVID-19 is new and finding an appropriate data set with enough details is difficult. Therefore, the data were collected from kaggle.com of x-ray images taken to patients' chests. The data is available publicly for Chinese patients from Dec. 2019 – Feb. 2020. There are 98 images taken from x-ray images.

Training Phase:

Sampling was random sampling and the repeat train/test is 10, while training set size 66%. The algorithms used for testing and benchmark were 4 algorithms SVM, Random Forest, Neural Network, and kNN. As preprocessing of data, the first step is to remove features with too many missing values with thresholds percentage is 5, second, select relevant features while the score is (information gain) with several features (fixed 10).

Results:

The receiver operating characteristics (ROC) curve is a performance measurement for classification problems at various threshold settings[18]. It is considered as a standard for evaluation and standard-setting, used for comparing False Positive Rate with True Positive Rate[19]. The confusion matrix is depicted in Fig. 1 and 2 where in these two figures there is a classification model with 4 different combinations of predicted and actual values, the light grey positive diagonal represents correct predictions, while the pink negative diagonal unfitting predictions.

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	1.0 %	240
	normal	0.0 %	99.0 %	100
	∑	239	101	340

(a) Confusion matrix for SVM

		Predicted		
		covid	normal	∑
Actual	covid	97.5 %	2.1 %	240
	normal	2.5 %	97.9 %	100
	∑	244	96	340

(b) Confusion matrix for Random Forest

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	1.0 %	240
	normal	0.0 %	99.0 %	100
	∑	239	101	340

(c) Confusion matrix for kNN

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	11.5 %	240
	normal	0.0 %	88.5 %	100
	∑	227	113	340

(d) Confusion matrix for Neural Network

Figure 1. Confusion matrix Before showing proportion of predicted

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	1.0 %	240
	normal	0.0 %	99.0 %	100
	∑	239	101	340

(a) Confusion matrix for SVM

		Predicted		
		covid	normal	∑
Actual	covid	98.8 %	1.0 %	240
	normal	1.2 %	99.0 %	100
	∑	242	98	340

(b) Confusion matrix for Random Forest

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	1.0 %	240
	normal	0.0 %	99.0 %	100
	∑	239	101	340

(c) Confusion matrix for kNN

		Predicted		
		covid	normal	∑
Actual	covid	100.0 %	11.5 %	240
	normal	0.0 %	88.5 %	100
	∑	227	113	340

(d) Confusion matrix for Neural Network

Figure 2. Confusion matrix After showing proportion of predicted

Baghdad Science Journal
2022, 19(6): 1356-1361
P-ISSN: 2078-8665
E-ISSN: 2411-7986

Categorizing the predicted positive or negative and the actual as true or false is done in this phase. The definition of sensitivity is the ratio of true positives to all positives. While specificity is defined as the ratio of true negatives to all negatives.

The positive and negative predicted values in addition to true and false actual values are categorized into different categories. The sensitivity is defined as the ratio of true positives to all positives. In the same way, specificity is defined as the ratio of true negatives to all negatives. Relying on the predicted and actual values and equations explained in reference[20], Eqs.1, 2 and 3. The mathematical relationships of sensitivity, specificity, and precision have been formulated using these three equations.

Sensitivity = TP / TP . FN 1

Specificity = TN / TN . FP 2

Precision = TP / TP . FP 3

In both Figs. 3, 4, the curve's AUCs is exactly 0.5, it represents the case prediction of the people diagnosed earlier as healthy and non-healthy. The high value of AUC means better diagnosing of the model can distinguish between the patients and healthy people.

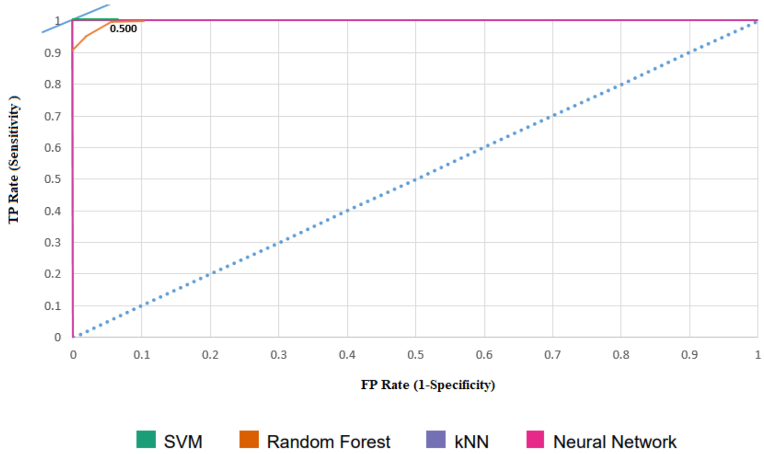

Figure 3. receiver operating characteristics (ROC) BEFORE

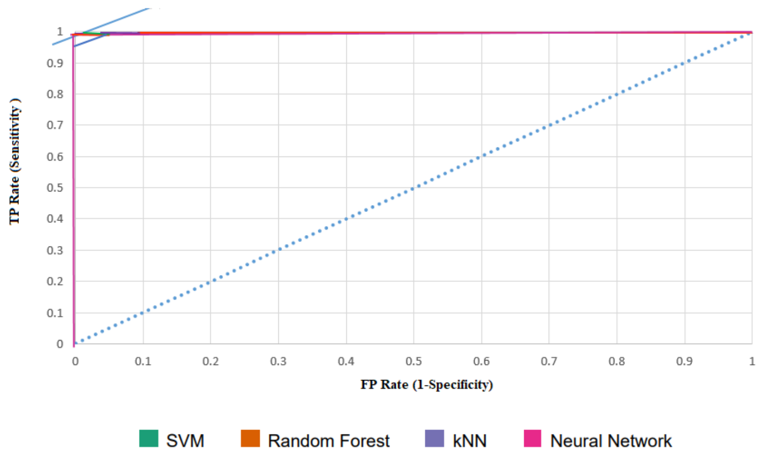

Figure 4. receiver operating characteristics (ROC) AFTER

Discussion:

A deep learning model for detecting COVID-19 cases in chest X-ray images was proposed in this research. The model was tested on the selected datasets and accomplished exceptionally well. The proposed model reached an accuracy of almost 98% classification tasks. An additional positive remark from the results is the precision and recall (Sensitivity) for COVID-19 cases. The high recall value means false negative (FN) cases are lower and this is considered a hopeful result. The importance of these findings means accurate diagnosing of the COVID-19 case which makes it possible to win the battle against this foe as this is the main aim of this research. Table 1 presents a performance comparison of the proposed model on the class classification task. ROC curves show that the PS model has a relative capacity to differentiate the positive from negative classes. Unlike the AT model, the PS model has a nonlinear performance which displays that it has values of sensitivity higher compared to AT model.

Baghdad Science Journal
2022, 19(6): 1356-1361
P-ISSN: 2078-8665
E-ISSN: 2411-7986

AI-based techniques for detecting COVID-19 can achieve high accuracy which gives an early move to control and eliminate the spread of the virus. This research achieved high accuracy in comparison with others as shown in Table 1. The high training with more realistic data makes the algorithm better in term of accuracy and successfulness which increase the reliability on AI-based techniques and reduce efforts on the human factor. In Table 1, the SVM (linear) overwrites the other eleven diagnostic models that have received a cabinet value from the ideal solution and the value furthest from the negative solution.

Table 1. Comparison of our proposed method with other benchmark techniques

Model	AUC	CA	F1	Precision	Recall
SVM	0.9999166666666667	0.9970588235294118	0.9970630697672145	0.997.879440885265	0.9970855235294118
Random Forest	0.9995625	0.9882352941176471	0.9882002845146609	0.988248191838719	0.9882352941176471
Neural Network	0.9995833333333334	0.961764705882353	0.962399327737851	0.9661634565330557	0.961764705882353
kNN	0.9978541666666667	0.9970588235294118	0.9970630697672145	0.9970879440885265	0.997058823594118

Conclusions:

Countries are facing COVID-19 with a shortage of resources and staff. This emergency needs intense effort and identifying of every single positive case accurately. Considering this emergency case, in this paper, a machine learning technique proposed to identify positive COVID-19 cases from chest X-ray images. The data set was retrieved from a public repository for Chinese patients. This type of data sets is used in other researches which proves its reliability. The proposed method is less expensive and promising with its accurate results. However, improving our work with huge data set with more training may give better results. This gives specialists in the medical field a deep understanding of the positive cases and helps them identify the true positives from negatives results. The results were satisfactory with the benchmarked techniques and this increases the reliability on AI-based techniques in identifying COVID-19 cases. Additional experimental research is necessary to verify the efficiency of the proposed technique using larger dataset containing higher number of COVID-19 cases.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- The author has signed an animal welfare statement.
- Ethical Clearance: The project was approved by the local ethical committee in Dijlah University College.

Authors' contributions statement:

S. M. Z. contribution to the manuscript is the Conception, design, interpretation and results discussion. M. M. J. worked on acquisition of data and results analysis & discussion. M. A. K. worked on revision and proofreading and results discussion.

References

1. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW XZ. Review of the clinical characteristics of coronavirus disease 2019 (covid-19). J Gen Intern Med. 2020;1(5).
2. WHO. World Health Orgnization. Report: Q&As on COVID-19 and related health topics. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019, September 2020.
3. Sun D, Li H, Lu X-X, Xiao H, Ren J, Zhang F-R LZ-S. Clinical features of severe pediatric patients with coronavirus disease 2019 in wuhan: a single center's observational study. World J Pediatr. 2020;1(9).
4. Ping An, Hongbin Chen, Xiaoda Jiang, Juan Su, Yong Xiao, Yijuan Ding, et al. Clinical Features of 2019 Novel Coronavirus Pneumonia Presented Gastrointestinal Symptoms But Without Fever Onset. http://dx.doi.org/102139/ssrn3532530. 2020
5. Al-Waisy AS, Mohammed1 MA, Al-Fahdawi S, Maashi MS, Garcia-Zapirain B, Abdulkareem KH, et al. COVID-DeepNet: Hybrid Multimodal Deep Learning System for Improving COVID-19 Pneumonia Detection in Chest X-ray Images. Comput Mater Contin. 2020;67(2):2409–29.
6. Al-Waisy A S, Al-Fahdawi S , Mohammed MA. COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Comput Springer. 2020.
7. Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang WJ. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology. 2020;292(2):E115–7.
8. Xingzhi Xie, MD • Zheng Zhong, MD • Wei Zhao, MD • Chao Zheng, MD • Fei Wang, MD • Jun Liu M. Chest CT for Typical Coronavirus Disease 2019

(COVID19) Pneumonia: Relationship to Negative RT-PCR Testing. Radiology. 2020;296(2):E41–5.
9. Adam Bernheim , Xueyan Mei, Mingqian Huang, Yang Yang, Zahi A Fayad, Ning Zhang, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology. 2020;295(2):685–691.
10. Lih OS, Jahmunah V, San TR, Ciaccio EJ, Yamakawa T, Tanabe M KM, et al. Comprehensive electrocardiographic diagnosis based on deep learning. Artif Intell Med. 2020;103:101789.
11. Wang X, Qian H, Ciaccio EJ, Lewis SK, Bhagat G, Green PH, et al. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. Comput Methods Programs Biomed. 2020;187:105236.
12. Dekhtiar J, Durupt A, Bricogne M, Eynard B, Rowson H KD. Deep learning for big data applications in CAD and PLM – Research review, opportunities and case study. Comput Ind. 2018;100:227–43.
13. Zaid Khalaf Hussien BND. Anomaly Detection Approach Based on Deep Neural Network and Dropout. Baghdad Sci J [Internet]. 2020;17: 701-709.
14. Hussain BA HM. Developing Arabic License Plate Recognition System Using Artificial Neural Network and Canny Edge Detection. Baghdad Sci J [Internet]. 2020;17(3): 909-915.
15. Emily Chen, Kristina Lerman EF. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Heal Surveill. 2020;6(2):e1927:1–12.
16. Joseph Paul Cohen, Paul Morrison LD. COVID-19 Image Data Collection. arXiv Prepr 200311597. 2020;1–4.
17. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA [Internet]. 2020 Apr 14;323(14):1406–7. https://doi.org/10.1001/jama.2020.2565
18. Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27:861–874.
19. Powers Dmw. Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation. J Mach Learn Technol [Internet]. 2011;2(1):37–63. Available from: http://www.bioinfo.in/contents.php?id=51
20. Gönen M. Analyzing Receiver Operating Characteristic Curves with SAS. J Biopharm Stat. 2008;18(6):1228–9.

تشخيص عدوى كوفيد-19 المستجد في صور الأشعة السينية باستخدام الشبكة العصبية

سالم محمد زكي [1] **مصطفى موسى جبر [2,3]** **محمد علاء الدين كشمولة [4]**

[1] قسم علوم الحاسوب، كلية دجلة الجامعة، بغداد، العراق
[2] قسم هندسة تقنيات الاجهزة الطبية، جامعة الفراهيدي، بغداد، العراق
[3] كلية التراث الجامعة، بغداد، العراق
[4] قسم علوم الحاسوب، كلية التربية، جامعة الحمدانية، نينوى، العراق

الخلاصة:
مع انتشاره السريع ، صدمت عدوى فيروس كورونا العالم وكان لها تأثير كبير على حياة المليارات من الناس .لقد ثبت أن صور الأشعة السينية هي طريقة مهمة لتحديد وتقدير ومراقبة الأمراض .يمكن استخدام خوارزميات التعلم العميق للمساعدة في تحليل أعداد ضخمة محتملة من فحوصات الأشعة السينية .لقد أجرينا نظام تحليل متعدد الاختبارات بأثر رجعي للكشف عن COVID-19 المشبوهة ، والأداء واستخدام ميزات الصور بالأشعة السينية لتقييم تقدم حالة المرض في كل مريض ، مما أدى إلى "اكتشاف وجود كورونا فايروس في الصور التي تم تحليلها وكانت النتائج مقبولة بالمقارنة مع التقنيات الاخرى. أظهرت نتائج هذا البحث أن التطور السريع لتحليل الصور المستند إلى الذكاء الاصطناعي (AI) يمكن أن يحقق دقة عالية في الكشف عن عدوى فيروس كورونا بالإضافة إلى القياس الكمي ومراقبة تطور حالة المرض.

الكلمات المفتاحية: كوفيد-19، التعلم العميق، المعالجة الصورية، الشبكات العصبية، معالجة صور أشعة-أكس