An accurate neural network algorithm to diagnose Covid-19 from CT images

H. Romdhane[1*], H. Dziri[1], M. Ali Cherni[2], D. Ben-Sellem[1,3,4]

[1]Université de Tunis El Manar, Laboratoire de recherche en Biophysique et Technologies Médicales (LRBTM), ISTMT
[2]Université de Tunis, LR13 ES03 SIME, ENSIT, Montfleury 1008 Tunisia
[3]Université de Tunis El Manar, Faculté de Médecine de Tunis, 1007, Tunis, Tunisia
[4]Institut Salah Azaiez, Service de Médecine Nucléaire, 1006, Tunis, Tunisia

▶ ***Original article***

***Corresponding authors:**
Hamida Romdhane, PhD.,
E-mail:
romdhane.hamida@gmail.com

Revised: October 2020
Accepted: November 2020

Int. J. Radiat. Res., April 2021; 19(2): 349-356

DOI: 10.29252/ijrr.19.2.349

ABSTRACT

Background: A new coronavirus appeared in late December 2019 in Wuhan, China. He was named Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). This virus is responsible for Covid-19, the name given to the disease associated with it. It spreads worldwide, infecting more than a million people and killing more than 70 miles. The rapid and accurate diagnosis of suspected Covid-19 cases plays a crucial role in medical treatment and timely quarantine. ***Materials and Methods:*** In order to counter the Covid-19 pandemic, we have developed a method for the automatic detection of Covid -19, from 2D computed tomography (CT) chest images. It is a supervised software system based on the ANN (Artificial Neural Network) algorithm. Pulmonary CT images were collected from multiple international datasets, with a total of 395 images: 70% were used for training and 30% were used for testing. For each patient, the lungs were segmented using simple thresholding. Then, the segmented lungs were fed into a neural network to predict the probability of SARS-CoV-2 infectious. ***Results:*** The internal validation achieved a total accuracy of 97.5% with a specificity of 96.6 % and a 100 % sensitivity. ***Conclusion:*** These results demonstrate the proof-of-principle for using artificial intelligence to extract radiological features for timely and accurate Covid-19 diagnosis.

Keywords: *Covid-19, chest ct images, SARS-CoV-2, neural network algorithm.*

INTRODUCTION

Since the end of December 2019, the world has been surprised by the appearance of an epidemic in Wuhan, which has spread to most countries of the world. This pandemic of Covid-19 is due to a new coronavirus, named by the World Health Organization SARS-CoV-2. The World Health Organization has declared a state of emergency against this pandemic (1, 2).

On April 6, 2020, 1,174,855 confirmed cases were infected by SARS-CoV-2 and 64,471 deaths in 209 countries (3). Fever, cough, dyspnea, myalgia, or fatigue is the main symptoms of the Covid-19 (1,4,5). The disease can progress in four stages (6). The first stage starts from 0 days to the fourth day after the onset of the initial symptom. Pulmonary computed tomography (CT) allows the detection of pulmonary anomalies from the second day (7). The second stage begins from the fifth day to the eighth day after the initial symptom onset. It is the progressive stage where the infection quickly worsened. Then, the peak stage starts from the ninth to the thirteenth day. It is characterized by the increase in the pulmonary areas affected by the new virus. After the fourteenth day, the absorption phase is started. This step is important for the patient's life and is critical for patients with chronic diseases such as diabetes and heart failure.

The standard diagnostic technique is the reverse transcription-polymerase chain reaction

(RT-PCR) method [8]. They are performed on clinical research samples of nasal secretions. These samples are collected by inserting a swab into the nostril. Then, it gently moves into the nasopharynx to collect secretions. Though, RT-PCR can identify the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains that cause Covid-19. In some instances, it produced negative test results even though the patients showed progression on follow-up chest computed tomography (CT) scans [9, 10]. The detection of Covid-19 symptoms in the lower parts of the lungs has a higher accuracy when using CT scans than when using RT-PCR [11]. In some cases, CT scans tests can be substituted with RT-PCR tests. However, they cannot exclusively address the problem due to the relatively limited number of radiologists, compared to new residents, and the high volume of re-examinations of infected people who wish to know their illness's progression.

Several studies [12, 13] have recommended using CT scans rather than RT-PCR due to its limited availability in some countries. To overcome the challenges of CT scans and assist radiologists, we need to improve the procedure's speed. This can be achieved by designing advanced diagnostic systems that utilize artificial intelligence (AI) tools.

The AI has shown its high performance in imaging lung diseases: evaluation of respiratory function in 2010 [14], automatic detection of pulmonary embolism in CT images in 2017 [15], and screening for lung cancer in 2019 [16]. Various researches have implemented AI to diagnose COVID-19 from CT scans. One of the main advantages of AI is that it can be implemented in a trained model to classify unseen images.

Xu *et al.* [17] reported that real-time RT-PCR has a low positive rate at the early phase of COVID-19. They developed a first screening model that uses deep-learning methods for differentiating COVID-19 pneumonia (viral pneumonia) from grippe and unchanging cases using pulmonary CT images. The dataset contains 618 CT samples. They were obtained for the analysis and they were classified as COVID-19, grippe (viral pneumonia), and other cases using ResNet-18 and ResNet-based methods. The authors used a noisy or Bayesian function to distinguish the infected images and obtained a detection accuracy of 86.7%.

Furthermore, Wang *et al.* [18] produced a deep -learning technique for extracting CT scans data. Their work included a collection of 99 patients (453 CT scans). They applied a modified network inception model and obtained an accuracy of specificity of 80.5% and a sensitivity of 84%.

Fu *et al.* [19] proposed a classification method based on ResNet-50 to detect COVID-19 and other infectious lung diseases (bacterial pneumonia and pulmonary tuberculosis). The authors collected a dataset of 918 patients (60,427 CT scans): 14,944 of CT scans were from 150 COVID-19 patients and 15,133 from 154 non-COVID-19 viral pneumonia patients. They performed various tests for many lung diseases. The achieved respectively, accuracy, sensitivity, and specificity were 98.8%, 98.2%, and 98.9%.

Meanwhile, Gozes *et al.* [20] employed a deep-learning approach to identify COVID-19 patients automatically. The dataset contains CT scans from 157 foreign patients (China and the USA). To evaluate their system, the authors applied Resnet-50-2 and obtained an area under the 99.6% curve. The sensitivity and specificity were 98.2% and 92.2%, respectively.

Jiang *et al.* [21] compared RT-PCR to CT scans and examined 51 patients (29 men and 22 women). The authors obtained a 98% sensitivity for the detection of COVID-19, compared to the initial RT-PCR sensitivity of 71%.

Thus, our work aims to develop automated chest image analysis tools based on a neural network and demonstrate their usefulness in differentiating coronavirus patients from those who do not have the disease to support the detection and monitoring of disease progression [22].

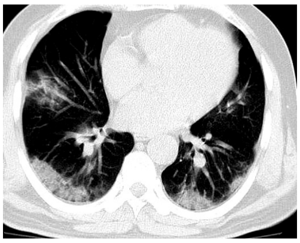
The paper is organized as follows: Section 1 presents the materials and methods used in this work. Section 2 presents our method's simulation results by offering a comparative study between the proposed method and other literature methods. In section 3, we discussed our results. Finally, we present the obtained results to the conclusion.

MATERIALS AND METHODS

Databases

This study was carried out on an international dataset including 395 images of pulmonary CT images: international public databases (Italy, Iran, Turky, Belgium and China) (23-25) containing 110 images of patients infected SARS-CoV-2, which was confirmed by positive laboratory tests. And a Tunisian database obtained from Salah Azaiez Institut having 285 images of 285 patients with no respiratory pathologies. The total number of male patients equal to 210 with age started from 25 to 85 years. Thus the age for 185 female patients varies from 30 to 85 years.90% of patients are presented with a fever and cough. Table 1 illustrates an example of the dataset.

Table 1. Examples of database.

Cases	Patients data	Diagnosis
	**AGE:** 70 years **GENDER:** Male **Symptom:** Recent travel from the endemic COVID-19 region. 24 hours of confusion with new temperature and desaturation on assessment.	Case contributed by Dr. Derek Smith (**Institution:** South East Scotland Radiology Training Scheme) Multifocal regions of consolidation and ground-glass opacifications. These have a peripheral and basal predominance. No pleural or pericardial effusion. Serial imaging demonstrating progressive changes in a patient with PCR confirmed COVID-19 infection. There can be a rapid deterioration in imaging findings.
	**AGE:** 55 years **GENDER:** Male **Symptom:** Fever and non-productive cough start from 5 days ago.	Case contributed by Dr. Bahman Rasuli (**Institution:** Jame Jam Imaging Center) Bilateral multi-lobar peripheral ground-glass and consolidative opacities are seen in both lungs, mostly mid to lower zones. Non-specific mediastinal lymph nodes. The RT-PCR COVID-19 assay was positive, so this patient was deemed to have COVID-19 pneumonia.

Classification methods

In the medical sector, artificial intelligence may have a significant effect on patient diagnosis and management. Several tools are used for classification; for this paper, we highlighted two of the well-known methods for classification: Support Vector Machines (SVM) and artificial neural networks (ANN).

Support Vector Machine

SVM is categorized from supervised machine learning, where algorithms can obtain a human-labeled database to predict the coveted and proven outcome. For example, suppose it is intended to predict whether a lung is infected with SARS-CoV-2. In that case, research should be carried out based on a safe dataset containing a collection of patients with such a feature and another collection not showing such abnormalities. Identifying a species through machine learning is key to identifying and recognizing the minute features or patterns within the input images. Therefore, the function extraction process effectively transforms the input images into their digital descriptors. For this study, Co-occurrence matrices (26) were adopted to extract features to train the SVM.

An SVM is a formally specified hyperplane-separating discriminative classifier. In other words, the algorithm outputs an optimal hyperplane that categorizes new instances, given labeled training data (supervised learning). The learning of the hyperplane is done by transforming the problem using the kernel functions.

The artificial neural networks

A neural network is a collection of algorithms that, through a mechanism that mimics the way the human brain works, aim to identify underlying relationships in a set of data. Neural networks, in this context, apply to neuron structures, either organic or artificial. Neural networks may adapt to evolving inputs, so the network produces the best possible outcome without the output parameters having to be revamped.

The artificial neural networks (ANN) algorithm is characterized by its ability to gradually improve its efficiency by exposing itself to large quantities of training data. It begins with random parameters in general. The outcomes of the first trials remain inadequate. To minimize a pre-defined error metric, this classifier is aimed at optimizing such parameters.

This is achieved by changing the ANN parameters to gradually reduce the error measurement for each data exposure iteration.

The proposed methodology

The methodology starts with a segmentation procedure to extract lungs from CT scans. This information was then introduced in the ANN algorithm to identify the COVID-19 patients from that non-COVID-19. The proposed method is described in figure 1.

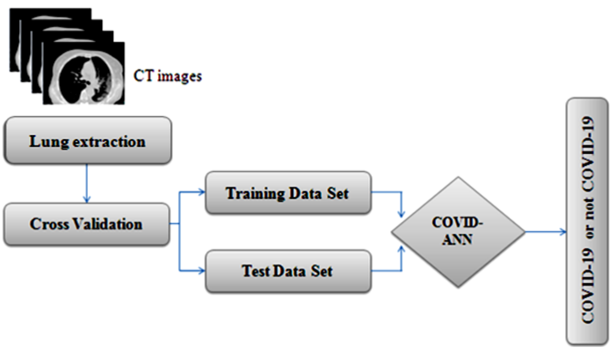


Figure 1. Classification Process for the COVID-19.

Identification of lung

The segmentation was verified in CT dataset to separate lungs from the background. Simple thresholding is applied to creating a lung map. Firstly, the thresholding process lets to extract the lungs from the background producing a binary image. Secondly, we adopt the morphological operation to find the perimeter of objects in this binary image. Finally, we display the results of detecting the lung's perimeter in the original image. We labeled regions, in which we display the data in the input image as an image that uses the full range of colors in the color map.

Hence, only the new images were collected for future classification steps. Figure 2 shows one case of covid-19 and one case of non-covid-19. The red circle indicates the area infected by the virus, which is characterized by ground-glass opacities and, therefore, by a lower density.

Artificial neural network structure

To solve a specific problem, an ANN is made up of highly interconnected processing neurons. It has three types of layers: the input layer, the hidden layers and the output layer. The input layer receives the input variables fed into the network. The hidden layer, comprised of several layers, sends data to the output layer. The latter presents the output. In our case, it differentiates Covid-19 cases from not covid-19 ones.

There are two groups of artificial neural networks: radial bases function and Multi-layer perception (MLP). In the present work, the proposed ANN named Covid-ANN is based on a backpropagation algorithm with a sigmoid exponential function and MLP structure.

The Covid-ANN model illustrated in figure 3 consists of an input layer, hidden layers and an output layer. The input layer is composed of 4096 neurons (64*64 pixels in the segmented image). Each neuron receives one pixel from the input image that fed forward it to 3 hidden layers. The output layer (one neuron) provides the input image's final result, classifying the patient as Covid-19 or not Covid-19.

We have used 70% of the dataset in the training step using cross-validation. Table 2 presents the main training parameters. As a transfer function for all layers, the log sigmoid [27] defined by the equation is used for the training phase.

$$F(x) = \frac{1}{(1+e^{(-x)})} \quad (1)$$

The gradient descent with momentum and ba ckpropagation of the adaptive learning rate trains the network. This function updates bias and weight values according to the momentum of gradient descent and an adaptive learning rate. The constant momentum (mc) was 0.8; the learning rate (LR) was 0.01, the maximum number of epochs was set at 100, and the minimum gradient is 10^{-7}.

The K-Fold Cross Validation technique was used to generalize classifier models and perfectly tested data by dividing the data into K classes, and Cross-Validation ensured that each fold from the K-Folds was used as a test collection. Using 10 K-Fold Cross-Validation, the ANN and SVM classifiers were trained.

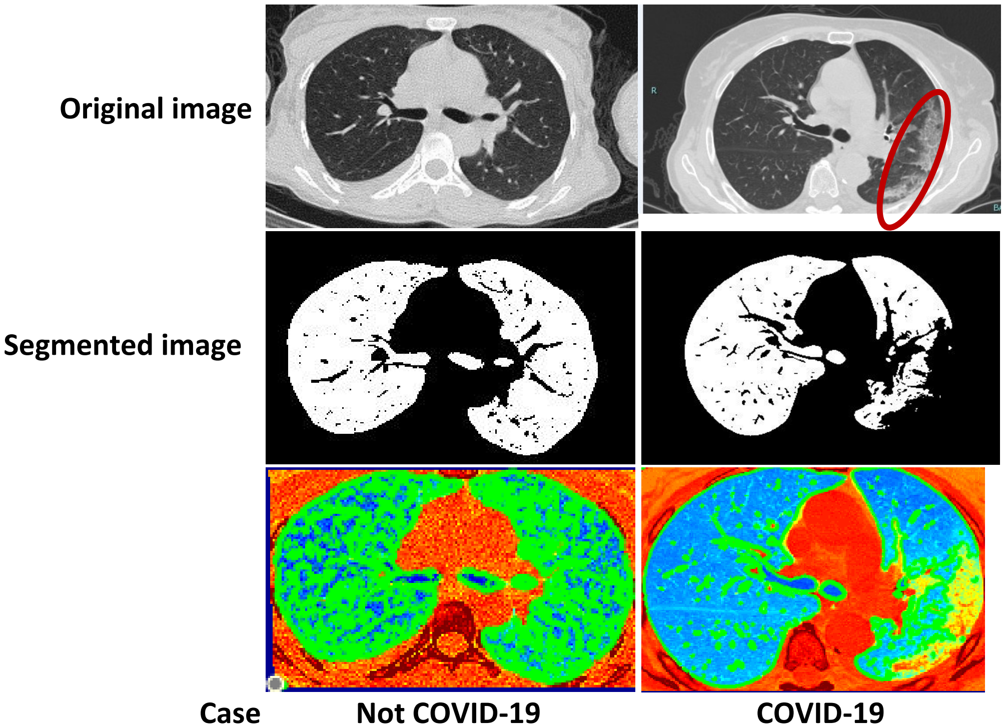


Figure 2. Lungs extracted from CT images (red circle indicates the area infected by the virus).

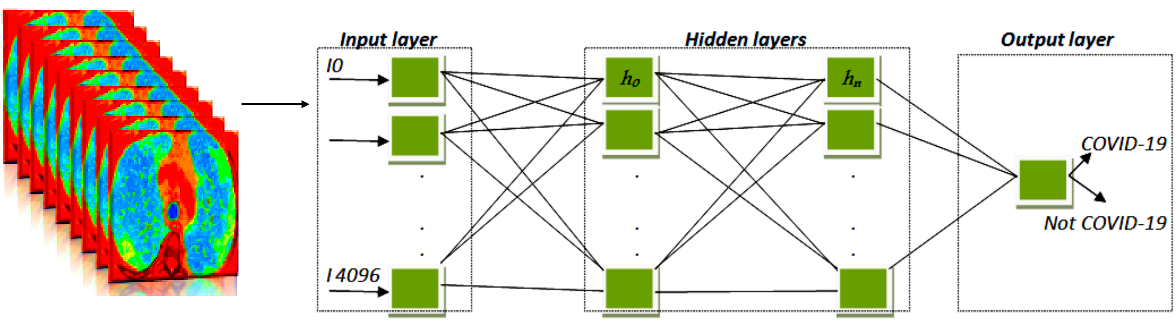


Figure 3. COVID-ANN architecture.

Table 2. Main training parameters.

Training Parameter	Parameter value
net.trainParam.epochs	100
net.trainParam.lr	0.01
net.trainParam.mc	0.8
net.trainParam.min_grad	10^{-7}

RESULTS

In this study, machine learning was applied to develop a classification network identifying COVID-19 patients. The test step was carried out in a total of 118 pulmonary CT images. The duration of the two steps (training and test) was short of 523,064 seconds.

We begin with an assessment of the ability of ANN to detect coronavirus patients versus non-corona virus patients. First, we compared the results obtained by both ANN and SVM to our patients' established diagnoses (table 3 and 4). Then, we compared the efficiency of ANN against the SVM algorithm (table 5).

Finally, to confirm our method's effectiveness, we compared to methods already found in the literature (table 6).

Table 3. Confusion matrix of SVM.

		Confirmed results		Total
		COVID-19	Not COVID-19	
COVID-SVM results	COVID-19	TP=27	FN=6	33
	Not COVID-19	FP=2	TN=83	85
Total		29	89	118

Table 4. Confusion matrix of ANN.

		Confirmed results		Total
		COVID-19	Not COVID-19	
COVID-ANN results	COVID-19	TP=30	FN=3	33
	Not COVID-19	FP=0	TN=85	85
Total		30	88	118

Table 5. Performance of the proposed artificial neural network.

Performance	Sensitivity	Specificity	Accuracy	AUC
SVM	93.10%	93.25%	93.23%	0.90
Proposed method	**100 %**	**96.6 %**	**97.5 %**	**0.95**

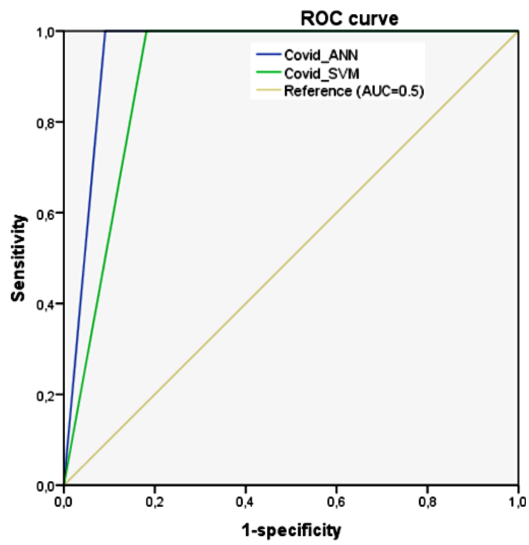
Table 6. Comparative table between the literature study and the proposed method.

	Accuracy	Specificity	Sensitivity	AUC	Method
Xu *et al.* (19)	86.7%.				ResNet-based methods
Wang *et al.* (18)	82.9%	80.5%	84%.		Deep-learning technique
Proposed method	**97.5%**	**96.6%**	**100%**	**0.95 %**	**ANN**

Tables 3 and 4 detail the results obtained for correctly identified cases (true-positive (TP) and true-negative (TN)) and those incorrectly classified (false-positive (FP) and false-negative (FN)).

A TP is a case identified Covid-19 by our algorithm Covid-ANN and which is a confirmed case. An FP is a case identified Covid-19 by our algorithm and which is an uninfected case. A TN is a case identified not Covid-19 by our algorithm and which is an uninfected case. FN is a case identified not Covid-19 by our algorithm and which is a confirmed case.

From these results, we can calculate some metrics that are sensitivity, specificity, accuracy (table 5) and AUC (Area under Curve) figure (4) [28].



- Sensitivity is the ability of a test to correctly classify an individual as 'diseased'

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

- Specificity is a test's ability to correctly classify an individual as *disease-free* is called the test's specificity.

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

- Accuracy is one of the criteria for evaluating classification models. Informally, it refers to the proportion of correct predictions made by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

With TP= number of true-positive
TN= number of true-negative
FP= number of false-positive
FN= number of false- negative

- AUC (area under the ROC (Receiver Operator Characteristic) curve): a combination of sensitivity and specificity can also be obtained by calculating the area under the ROC curve (AUC). The larger this surface (AUC max = 1), the more reliable the automated diagnostic system.

DISCUSSION

Classification results for Coronavirus vs. Non-coronavirus cases per the CT studies were 0.95 AUC, 100 % sensitivity and 96.6 % specificity.

We obtain an accuracy value equal to 0.97, or 97.5 %. It could be concluded that our COVID-ANN very satisfactory results from the detection of the non-COVID patient.

Firstly, we will show the efficiency of the chosen classification method ANN with another classification algorithm, which is the SVM. Table 5 shows that ANN has high values of sensitivity, specificity, and accuracy than the values obtained by the SVM algorithm.

The COVID-ANN presents a high AUC equal to 0.95 compared by 0.90 achieved by the SVM.

Secondly, we compared our method to some method that exists in the literature. As shown in Table 6, our method achieved higher values of sensitivity compared to Xu *et al.* [18] (ResNet-based methods). Also, it achieved the higher values of sensitivity, specificity, and accuracy than the values obtained by Wang *et al.* [17] (Deep-learning technique).

These values indicate that the detection of Covid-19 by the proposed method was very close to the detection by an expert. This also proved the effectiveness of the proposed method in the detection of Covid-19.

The results demonstrate that our proposed method can automatically extract the presence of a region infected by the Covid-19 and achieve good performance. This is the main principle benefit of the neural network that can solve various types of problems and extract the important features for accurate classification.

CONCLUSION

In conclusion, we proposed an automatic system for detecting lung lesions caused by the new coronavirus SARS-CoV-2, which attacks lung cells and caused Covid-19. We have developed a robust automatic machine learning model based on an artificial neural network using a multi-layers networks model, which diagnoses a high accuracy Covid-19 patients from thoracic CT images.

Conflicts of interest*: *Declared none.

REFERENCES

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cheng Z (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, ***395(10223)**: 497-506*.
2. Lu H, Stratton CW, Tang YW (2020) Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of Medical Virology*, ***92(4)**: 401-402*.
3. Coronavirus, WHO (2020) URL https://www. WHO. int/ emergencies/diseases/novel-coronavirus-2019. *Library Catalog: www. WHO, int*.
4. Liu T, Huang P, Liu H, Huang L, Lei M, Xu W, Liu B (2020) Spectrum of chest CT findings in a familial cluster of Covid-19 infection. *Radiology: Cardiothoracic Imaging*, ***2(1)**: e200025*.
5. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Yu T (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, ***395(10223)**: 507-513*.
6. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng C (2020) Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (covid-19) pneumonia. *Radiology, 2020 Jun; **295(3):** 715-721*
7. Cecilia M, Rita RM, Cristina P, Filippo T, Luciano A, Stefano N, Francesco M (2020) Differential Diagnosis in Patients with HRCT Patterns Suspected for Covid-19 Pneumonia and RT-PCR Negative: A Multidisciplinary Approach for an Appropriate Management.
8. World Health Organization (2020) *Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human*

cases: interim guidance, 2 March 2020 (No. WHO/COVID-19/laboratory/2020.4). World Health Organization.

9. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Xia L (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (Covid-19) in China: a report of 1014 cases. *Radiology*, 200642.
10. Why can Covid-19 fatality in space be significantly higher than on earth (2020). *Int J Radiat Res, **18(3)**: 421-426*.
11. Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*.
12. Maghdid HS, Asaad AT, Ghafoor KZ, Sadiq AS, Khan MK (2020) Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. *arXiv preprint arXiv:2004.00038*.
13. Bukhari SUK, Bukhari SSK, Syed A, SHAH SSH (2020) The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with Covid-19. doi: https://doi.org/10.1101/2020.03.26.20044610
14. Ozkan H, Tulum G, Osman O, Sahin S (2017) Automatic detection of pulmonary embolism in CTA images using machine learning. *Elektronika ir Elektrotechnika*, ***23(1)**: 63-67*.
15. Sannasi Chakravarthy SR and Rajaguru H (2019) Lung cancer detection using probabilistic neural network with modified crow-search algorithm. *Asian Pacific Journal of Cancer Prevention: APJCP*, ***20*(7)**: *2159*.
16. Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Internal Medicine*, ***4(2)***: *627*.
17. Butt C, Gill J, Chun D, Babu BA (2020) Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, ***1**: 1563-72*. A Journal of Composition Theory
18. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Xu B (2020) A deep learning algorithm using CT images to screen for Corona virus disease (Covid-19). Eur Radiol (2021). https://doi.org/10.1007/s00330-021-07715-1
19. Fu M, Yi SL, Zeng Y, Ye F, Li Y, Dong X, Zhang Q (2020) Deep learning-based recognizing Covid-19 and other common infectious diseases of the lung by chest CT scan images. doi: https://doi.org/10.1101/2020.03.28.20046045
20. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, Siegel E (2020) Rapid AI development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. *arXiv preprint arXiv: 2003.05037*.
21. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z (2020) Review of the clinical characteristics of coronavirus disease 2019 (Covid-19). *Journal of General Internal Medicine*, *1-5*.
22. Jafari S, Arabalibeik H, Agin K (2010, April) Classification of normal and abnormal respiration patterns using flow volume curve and neural network. In *2010 5th International Symposium on Health Informatics and Bioinformatics* (pp. 110-113). IEEE.
23. Covid-19 pneumonia: https://radiopaedia.org/cases/covid-19-pneumonia-8?lang=us
24. Covid-19 database: https://www.sirm.org/category/senza-categoria/covid-19/
25. Covid-chestxray-dataset: https://github.com/ieee8023/covid-chestxray-dataset
26. Xiao F, Kaiyuan L, Qi W, Yao Z, Xi Z (2018) Texture analysis based on gray level co-occurrence matrix and its application in fault detection. In *International Geophysical Conference, Beijing, China, 24-27 April 2018* (pp. 836-839). Society of Exploration Geophysicists and Chinese Petroleum Society.
27. Li B, Wang Y, Wang Y, Chen Y, Yang H (2014) Training itself: Mixed-signal training acceleration for memristor-based neural network. In *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)* (pp. 361-366). IEEE.
28. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R (2008) Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, ***56(1)**: 45*.