DOI: 10.24996/ijs.2021.62.8.27

ISSN: 0067-2904

Medical Image Classification for Coronavirus Disease (COVID-19) Using Convolutional Neural Networks

Maad M. Mijwil[1*], Ehsan Ali Al-Zubaidi[2]
[1]Baghdad College of Economic Sciences University, IRAQ
[2]Ehsan Ali Al-Zubaidi, University of Kufa, IRAQ

Received: 18/7/2020 Accepted: 23/10/2020

Abstract

The coronavirus is a family of viruses that cause different dangerous diseases that lead to death. Two types of this virus have been previously found: SARS-CoV, which causes a severe respiratory syndrome, and MERS-CoV, which causes a respiratory syndrome in the Middle East. The latest coronavirus, originated in the Chinese city of Wuhan, is known as the COVID-19 pandemic. It is a new kind of coronavirus that can harm people and was first discovered in Dec. 2019. According to the statistics of the World Health Organization (WHO), the number of people infected with this serious disease has reached more than seven million people from all over the world. In Iraq, the number of people infected has reached more than twenty-two thousand people until April 2020. In this article, we have applied convolutional neural networks (ConvNets) for the detection of the accuracy of computed tomography (CT) coronavirus images that assist medical staffs in hospitals on categorization chest CT-coronavirus images at an early stage. The ConvNets are able to automatically learn and extract features from the medical image dataset. The objective of this study is to train the GoogleNet ConvNet architecture, using the COVID-CT dataset, to classify 425 CT-coronavirus images. The experimental results show that the validation accuracy of GoogleNet in training the dataset is 82.14% with an elapsed time of 74 minutes and 37 seconds.

Keyword: CT- coronavirus image, GoogleNet, COVID-19, Deep learning, Convolutional neural networks.

تصنيف الصور الطبية لمرض فيروس كورونا (كوفيد–19) باستخدام الشبكات العصبية التلافيفية

معد محسن مجول[1*]، احسان علي كريم[2]
[1]كلية بغداد للعلوم الاقتصادية الجامعة، بغداد، العراق
[2]كلية التخطيط العمراني، جامعة الكوفة، النجف الاشرف، العراق

الخلاصة

الفيروس التاجي هو عائلة من الفيروسات التي تسبب أمراض خطيرة مختلفة تؤدي إلى الموت. تم العثور على نوعين من هذا الفيروس سابقًا: سارس SARS-CoV ، الذي يسبب متلازمة تنفسية حادة ، و -MERS CoV ، الذي يسبب متلازمة تنفسية في الشرق الأوسط. يُعرف أحدث فيروسات التاجية التي نشأت في مدينة ووهان الصينية باسم جائحة COVID-19. إنه نوع جديد من الفيروسات التاجية التي يمكن أن تؤذي الناس وتم اكتشافه لأول مرة في ديسمبر 2019. وفقًا لإحصاءات منظمة الصحة العالمية (WHO)، وصل عدد

*Email: mr.maad.alnaimiy@baghdadcollege.edu.iq

الأشخاص المصابين بهذا المرض الخطير إلى أكثر من سبعة ملايين شخص من في جميع أنحاء العالم وصل عدد المصابين في العراق إلى أكثر من 22 ألف شخص لغاية شهر نيسان 2020. في هذه المقالة، قمنا بتطبيق الشبكات العصبية التلافيفية (ConvNets) من أجل دقة الكشف عن صور التصوير التاجي المقطعي المحوسب (CT) التي تساعد الطاقم الطبي في المستشفيات على تصنيف صورة CT– coronavirus في مرحلة مبكرة. إن ConvNets قادرة على التعلم واستخراج الميزات تلقائيًا من مجموعة بيانات الصور الطبية. الهدف من هذه الدراسة هو تدريب بنية GoogleNet ConvNet باستخدام مجموعة بيانات COVID–CT لتصنيف 425 صورة CT–coronavirus. أظهرت النتائج التجريبية أن دقة شبكة GoogleNet هي 82.14٪ في تدريب مجموعة البيانات مع الوقت المنقضي هو 74 دقيقة و 37 ثانية.

1. Introduction

The COVID-19 virus belongs to the coronavirus (CoV) subfamily [1, 2] of the viruses family, and specifically, to the beta genus (beta-coronavirus). The SARS-CoV, that caused the outbreak that emerged in China in 2003 [3] [4], and MERS-CoV, that caused the outbreak that appeared in the Arabian Peninsula in 2012 [5], are of zoonotic origin (the original host is the bat), which at one point have evolved and crossed the barrier between species until causing the pandemic [6]. It should be noted that there are other coronaviruses that affect humans on a daily basis, causing mild diseases of the respiratory system, such as common colds (229E, NL63, OC43, or HKU1), all of which are from the alpha-coronavirus subgroup [7].

They receive this name (corona) for the structure they possess which, as seen under a microscope, gives them a kind of crown on their outer surface [8]. Viruses consist essentially of genetic material and structural proteins that encapsulate it. Figure-1 presents the structure of COVID-19 [9]. They consist of the nucleocapsid, where the genetic material is strictly contained (exclusively a simple ribonucleic acid (RNA) sequence of around 32,000 bases) and packaged thanks to protein N, and the envelope which is made up of various structural proteins such as the membrane glycoprotein or M protein, involved in virus assembly and in contact with the nucleocapsid, protein S, which forms the spikes responsible for adhesion to the host cell, and protein E, which interacts with protein M for the formation of the envelope [10, 11]. These represent the most relevant structural proteins, although others are vital for the replicative process of the virus as well as infection and entry into cells. It can be observed that the genomic sequence of COVID-19 is relatively similar to that of SARS-CoV (79%) and somewhat more different from that of MERS-CoV (50%) [12, 13]. COVID-19 is characterized by having a relatively high mutation capacity compared to other viruses, which makes the development of specific diagnostic methods, therapies, and vaccines somewhat complex.

The pandemic is still spreading with high speed and afflicts a huge number of people of different ages and genders, with no medicine developed. All institutions, schools, and universities have been closed and curfew and quarantine measures have been imposed in many cities all over the world. Tourism and travelling abroad have been banned. The critical question is when will a medicine be produced for this pandemic that hits the planet.

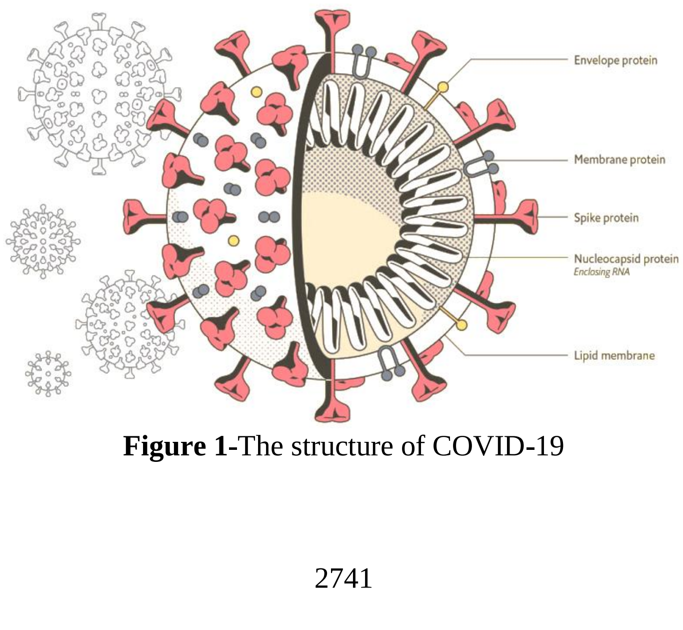

Figure 1-The structure of COVID-19

Very immediately, many studies appeared in 2020 about the use of deep learning techniques to classify a set of coronavirus images. The study conducted by Butt *et al*. [14] used multiple CNN models to classify and compare CT images of infections with coronavirus, influenza virus, pneumonia, and non-infection. The results indicated 98.2% sensitivity and 92.2% specificity. Narin *et al*. [15] suggested automatic CNNs-based transfer models (InceptionV3, Inception-ResNetV2, and ResNet50) for the prediction of COVID-19 in chest Xray images. The ResNet50 model had the best accuracy result for training, which is 98%. Singh *et al*. [16] proposed a multi-objective differential evolution–based CNNs models for classifying chest CT images for COVID-19 patients. The results showed accuracy of 1.9789%, F-measure of 2.0928%, sensitivity of 1.8262%, specificity of 1.6827%, and Kappa statistics of 1.9276%. Ozturka *et al*. [17] proposed DarkNet model as a classifier for YOLO real-time object detection system, with 17 convolutional layers and several filtering steps in each layer. This study obtained accuracy values of 98.08% and 87.02% for binary classes and multi-class cases, respectively. Amyar *et al*. [18] proposed a multi-task deep learning for the classification and segmentation of CT COVID- 19 images. The study included a dataset of more than 10000 patients (449 patients with the pandemic, 100 not infected, 98 with lung cancer, and 397 of other cases). The results showed a Dice coefficient value of < 0.78 for the segmentation and an area under the ROC curve value of < 93% for the classification. Another investigation [19] proposed deep learning and support vector machine (SVM) to detect COVID-19 disease. The results, based on the X-ray images, achieved accuracy = 95.33%, sensitivity= 95.33%, FPR= 2.33%, F1= 95.34%.

2. Chest CT- COVID-19 Medical Image

This test can assist to diagnose and monitor diseases, such as lung cancer, heart failure, sarcoidosis, pneumonia, tuberculosis, lung fibrosis, and COVID-19 disease. Physicians use X-rays to find out if treatments are effective and to check for any complications after a procedure or surgery. The following figures demonstrate infections with COVID-19 for people of different ages.

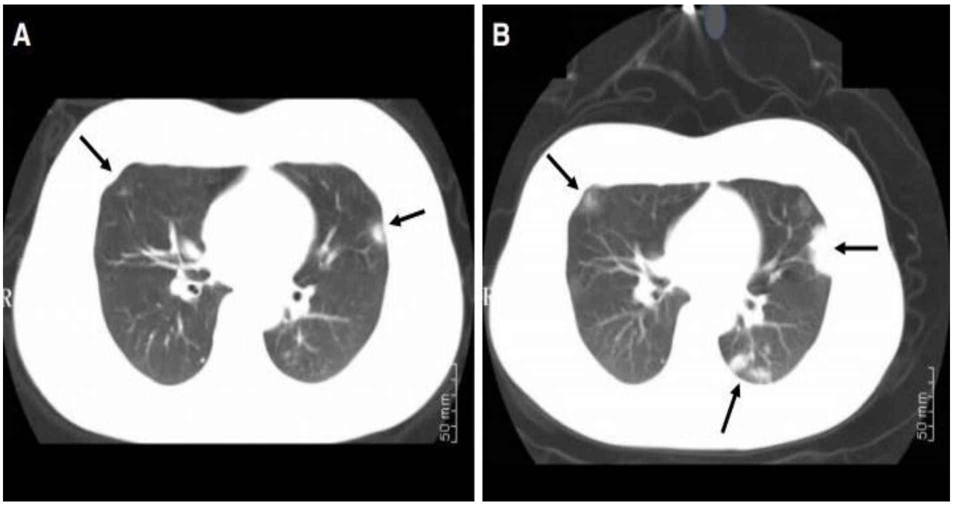

Figure 2- Example of CT-coronavirus medical image for a 27yr old woman.

Figure-2 [20] shows a coronavirus infection for a 27yr old woman, 36 weeks after the infection; A: The CT scan describes the characteristic peripheral and subpleural ground-glass opacities. These are observed in the left lower lobe/lingula junction and the right middle lobe (see arrow); B: The density, size, and distribution of these opacities have progressed (see arrow) after 2 days of confirmation.

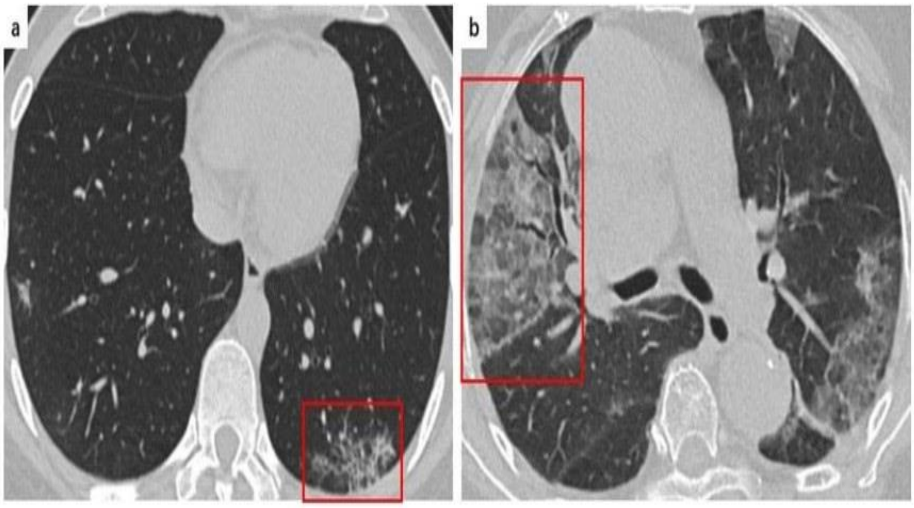

Figure 3- Example of CT-coronavirus medical image for a 35yr old man (A) and a 47yr old man (B).

Figure-3 [21] shows a coronavirus infection CT image for A: A man (35yr) with COVID-19 who suffers from severe headaches and fever after one day. CT-scan explains pure ground-glass opacity in the right lower lobe (red frame). B: A man (47yr) with COVID-19 patient who suffers from fever after seven days. CT-scan displays consolidation in the right lobe subpleural area (red frame).

3. The Proposed Method

The proposed method includes 3 steps (pre-processing, retraining GoogleNet ConvNet, and classification). Figure-4 presents the steps of the proposed method, which are also described below.

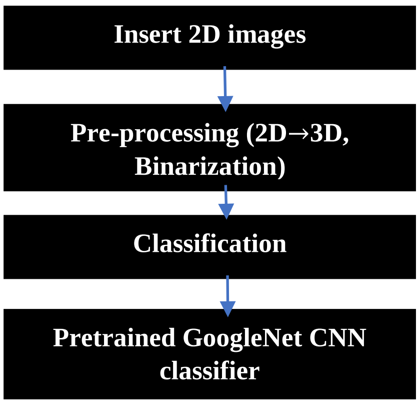

Figure 4-The steps of the proposed method in analysis of CT- scan images

1st step: The CT-coronavirus image contains a set of unnecessary information when it is taken. In order to obtain an adequate and perfect image, the recorded CT-coronavirus images must be pre-processed.
2nd step: In this step, some operations are executed on a CT-coronavirus image, e.g. the enhancement, normalization, noise reduction, filtering, binarization, and thinning. The binarization process is applied since it is the most important pre-processing step that is applied before retaining the image dataset. Besides, this step includes converting the image from 2D grayscale to 3D colour as well as saving these images.
The goal of the conversion to three-dimensional images is that CNNs accept this type of images. The Matlab function was applied for this process, as follows:

$$augmented\_set\_training = augmentedData\ Image\ store\ (Size\ image, Set\ training, '\ Preproces.Color', 'Grey2RGB');$$

$$augmented\_set\_test = augmented\ Data\ Image\ store\ (Size\ image, Set\ test, '\ Preproces.Color', 'Gray\ 2\ RGB');$$

In the binarization process, the $imbinarize$ function is applied to create a binary image of grayscale (2D)/colour (3D). This function is performed by replacing all values above a globally determined threshold (S) with $1S$ and setting all other values to $0S$. The $imbinarize$ function uses Otsu's segmentation method, which determines the S value to minimize the intraclass variance of the S black and white pixels. $imbinarize$ utilizes a 256-bin image histogram to compute Otsu's threshold. Applying the Otsu segmentation method, a robust S is computed based on the enhanced histogram. The best S is obtained using an exhaustive search based on all likely S values from 0 to 255.

3rd Step: Classification according to GoogleNet ConvNet architecture to retrain CT-dataset achieves high accuracy and reduces computing costs. GoogleNet is identified as InceptionV1. It presents the new theory of the inception block in ConvNet, in which multiscale convolutional transformations are integrated using split, transformation, and merge ideas. This block encapsulates filters of various sizes ($onexone, threexthree$ & $fivexfive$) to obtain spatial information at various scales. Figure-5 shows the sample images that have been used in this implementation.

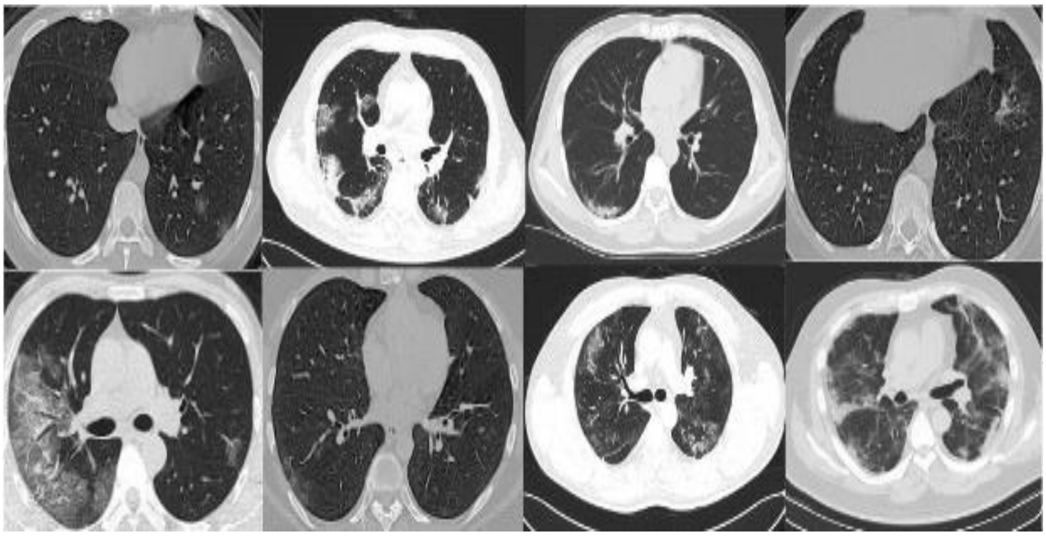

Figure 5-Sample of CT-dataset applied for retraining GoogleNet (from Google images).

Then, convolutional layers are replaced in tiny blocks, as in the idea of replacing each layer with micro-NN, as suggested in the NIN (Network in Network) architecture. GoogleNet's use of the idea of a split, transform, and merge assists solving a problem associated with learning different types of variations in the same class of several images. Besides, to increase learning capacity, GoogleNet focuses on making ConvNet parameters useful.

4th step: After retaining GoogleNet, each newly entered COVID-CT image can be classified.

4. Experiment Results

ConvNets are one of the deep learning methods applied in image classification, detection, and object recognition. The essential ConvNet architecture involves the convolution layer, pooling layer, and fully connected layer. All images obtained from the Google Images are unmodified and non-duplicate, as presented in Figure-5. Actually, it was a very difficult and tedious task to collect these images in high-resolution. There are not many images related to this topic on the Google site. Most of the studies are recent, where a pre-print version is also available on the Medrxiv and ResearchGate sites. All of these images are uploaded to the authors' profiles on the Mendeley website and can be used and shared by everyone. The proposed model execution runs by MATLAB v.2018 with a computer that has specifications of GPU=GeForce GTX 1070, CPU= Core i7 (Intel), RAM=16 GB, and Win=10 home. The dataset includes 425 CT-COVID images. Figure-6 shows the results of classification by GoogleNet ConvNet.

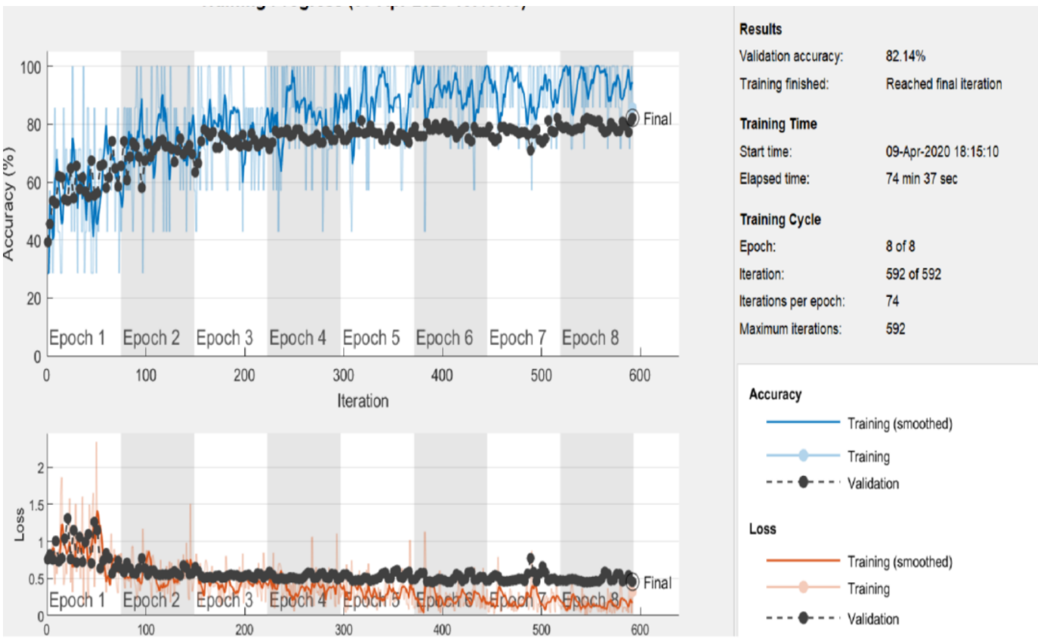

Figure 6- Results of images classification by GoogleNet CNN CT-Dataset.

Table-1 presents the present results of the effects obtained by the classification, in terms of the values of accuracy, sensitivity, specificity, and precision. ReLu is an activation function that is applied in all layers. To explain the metrics used when applying the classifier and obtaining results, it is necessary to define some concepts:

Accuracy: percentage of CT-scan images classified correctly. That is, we evaluate the percentage of success that the network has in our test set, as shown in equation 1:

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Sensitivity: percentage of success in cases of patients with covid-19, as shown in equation 2:

$$\frac{TP}{TP + FN} \tag{2}$$

Specificity: percentage of success in cases of healthy patients or with traditional pneumonia, as shown in equation 3:

$$\frac{TN}{TN + FP} \tag{3}$$

Precision: percentage of cases predicted by the model as coronaviruses that are actually coronaviruses, as shown in equation 3:

$$\frac{TP}{TP + FP} \tag{4}$$

True-positive (TP) denotes the cases of COVID-19 that are also classified by the model as positive. True-negative (TN)denotes cases that are not COVID-19 that are also classified by the model as negative. False-positive (FP)denotes the cases that are not COVID-19 but are classified by the model as positive. False-negative (FN)denotes the cases of coronavirus that are classified by the model as negative.

Table 1-The results of the CT- images classification accuracy, sensitivity, specificity, and precision.

Accuracy	Sensitivity	Specificity	Precision
84.14	83.12	82.11	84.01

The details in Figure-6 show the following: The validation accuracy is 82.14%. Elapsed time is 74 min and 37 sec (The more layers, the more accurate the results are, with 23 layers being applied in this paper). No. of iterations is 592. No. of epochs is 8. In addition, no. of iterations per epoch is 74.

The training opportunities in the research were as in the following:
$options = trainingOptions('sgdm', \ldots$
$'MiniBatchSize', 7, \ldots$
$'MaxEpochs', 8, \ldots$
$'InitialLearnRate', 1e-4, \ldots$
$'ValidationData', augmentedTestSet, \ldots$
$'ValidationFrequency', 3, \ldots$
$'ValidationPatience', Inf, \ldots$
$'Verbose', true, \ldots$
$'Plots', 'training-progress');$

5. Conclusions and Future works

Artificial intelligence techniques play an important and effective role in uncovering the COVID-19 pandemic. Specialists employ artificial intelligence techniques to examine this virus, test possible treatments, diagnose infected and uninfected people, and analyse the effects on public health. In this paper, we apply a deep learning technique (GoogleNet ConvNet) for the classification of a set of CT-coronavirus images (425 images) with InceptionV1 model. The experimental results confirmed that the validation accuracy of this test is 82.14%. In addition, this model achieved a precision of 84.01%, sensitivity of 83.12%, and specificity of 82.11%. To the best of our knowledge, this study is the first to use this approach. There shall be other studies in the future that deal with the application of artificial intelligence techniques in classifying and detecting the X-ray and/or CT-scan of COVID-19 pandemic and comparing them with images of viral pneumonia and bacteria.

References

1. Chhikara B. S., Rathi B., Singh J., and Poonam. **2020**. "Corona virus SARS-CoV-2 disease COVID-19: Infection, prevention and clinical advances of the prospective chemical drug therapeutics," *Chemical Biology Letters*, **7**(1): 63-72, March 2020.
2. Dhama K., Sharun K., Tiwari R., Sircar S., Bhat S., Malik Y. S., Singh K. P., Chaicumpa W., Bonilla-Aldana D. K., Morales A. J. R. **2019**. "Coronavirus Disease 2019– COVID-19," *Preprints*, 1-75, April 2020, https://doi.org/10.20944/preprints202003.0001.v2.
3. Yanga Y., Peng F., Wang R., Guana K., Jiang T., Xu G., Sun J., Chang C. **2020**. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China, Journal of Autoimmunity, 1-7, https://doi.org/10.1016/j.jaut.2020.102434.
4. Mijwil, M. M. **2021**. "Implementation of Machine Learning Techniques for the Classification of Lung X-Ray Images Used to Detect COVID-19 in Humans," *Iraqi Journal of Science*, **62**(6): 2099-2109, July 2021. https://doi.org/10.24996/ijs.2021.62.6.35.
5. Ahasan H. N., Das A., Chowdhury M. K. **2014**. "Middle East Respiratory Syndrome Coronavirus (MERS CoV): An Emerging Pathogen," *Journal of Medicine*, **14**(2): 156-163, July 2014, https:// doi. org/10.3329/jom.v14i2.19634.
6. Lai C., Shih T., Ko W., Tan H., Hsueh P. **2020.** "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *International Journal of Antimicrobial Agents*, **55**: 1-9, https://doi.org/10. 1016 /j.ijantimicag.20 20 . 105924
7. Liu D. X., Liang J. Q., and Fung T. S. **2020**. Human Coronavirus-229E, -OC43, -NL63, and -HKU1, *Elsevier Special Issue*, May 2020, https://doi.org/10.1016/B978-0-12-809633-8.21501-X.
8. Gorbalenya A. E., Baker S. C.,Baric R. S., de Groot R. J., Drosten C., Gulyaeva A. A., Bart L. Haagmans B. L., Lauber C., Leontovich A. M., Neuman B. W., Penzar D., Perlman S., Poon L. L. M., Samborskiy D. M., Sidorov L. A., Sola I., Ziebuhr J. **2020.** "The species Severe acute respiratory syndromerelated coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," *Nature Microbiology*, **5**(5): 1-9, https://doi.org/10.1038/s41564-020-0695-z.
9. Coronavirus Article April **2020**, Editors of Encyclopaedia Britannica, Available online (PDF file): britannica.com/science/coronavirus-virus-group,
10. Anastasopoulou S., Mouzaki A. **2020**. "The biology of SARS-CoV-2 and the ensuing COVID-19," *Achaiki Iatriki*, 39(1):29–35, April 2020.

11. Zhang X., Jia R., Shen H., Wang M., Yin Z., and Cheng A. **2017**. "Structures and Functions of the Envelope Glycoprotein in Flavivirus Infections," Viruses, **9**(11): 1-14, November 2017, https://doi .org /10.3390/v9110338.
12. Anand K. B., Karade S., Sen S., Gupta R.M. **2020**. SARS-CoV-2: Camazotz's Curse, *Medical Journal Armed Forces India*, **76**(2): 136-141, April 2020, https://doi.org/ 10.1016 /j.mjafi .2020. 04.008
13. Nassar M. S., Bakhrebah M. A., MEO S. A., Alsuabeyl M. S., W.A. Zaher W. A. **2018**. "Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: epidemiology, pathogenesis and clinical characteristics," *European Review for Medical and Pharmacological Sciences*, **22**: 4956-4961, 2018.
14. Butt C., Gill J., Chun D., Babu B. A. **2020**. "Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia," *Applied Intelligence*,1-7, https://doi.org/10.1007/s10489-020-01714-3
15. Narin A., Kaya C., Pamuk Z. **2020**. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks, *Arxiv*, March 2020.
16. Singh D., Vaishali V. K., Kaur M. **2020**. "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks," *European Journal of Clinical Microbiology & Infectious Diseases*, March 2020, https://doi.org/ 10.1007/s10096-020-03901-z
17. Ozturka T., Talo M., Yildirim E. A., Baloglu U. B., Yildirim O., Acharya U. R. **2020**. "Automated Detection of COVID-19 Cases Using Deep Neural Networks with X-ray Images," Computers in Biology and Medicine, **121**: 1-26, April 2020, https://doi.org/10.1016 /j. compbiomed.2020.103792
18. Amyar A., Modzelewski R., Ruan S. **2020**. "Multi-Task Deep Learning Based Ct Imaging Analysis for Covid-19: Classification And Segmentation," *MedRxiv*, April 2020, https://doi.org /10. 1101/2020.04.16.20064709
19. Sethy P. K., Behera S. K., Ratha P. K., Biswas P. **2020**."Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine," *Preprints.org*, April 2020, https://10.33889/IJMEMS.2020.5.4.052.
20. Chen R., Chen J., Meng Q. **2020**. "Chest computed tomography images of early coronavirus disease (COVID-19),"*Canadian Journal of Anesthesia*, **67**: 754–755, March 2020, https://doi.org /10.1007/s12630-020-01625-4
21. Ye Z., Zhang Y., Wang Y., Huang Z., Song B. **2020**. "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review," *European Radiology*, March 2020, https://doi.org/10.1007/s00330-020-06801-0.