CoviDecode: Detection of COVID-19 from Chest X-Ray images using Convolutional Neural Networks

Rishabh Raj

Abstract

Deep learning plays a major role from past years in image detection, spam re-organization, normal speech command, product recommendation and medical diagnosis. The detection of severe acute respiratory syndrome corona virus 2 (SARS CoV-2), which is responsible for corona virus disease 2019 (COVID-19), using chest X-ray images has life-saving importance for both patients and doctors. In addition, in countries that are unable to purchase laboratory kits for testing, this becomes even more vital. In this study, we aimed to present the use of deep learning for the high-accuracy detection of COVID-19 using chest X-ray images. Publicly available X-ray images were used in the experiments, which involved the training of deep learning and machine learning classifiers. Experiments were performed using convolutional neural networks and machine learning models. Images and statistical data were considered separately in the experiments to evaluate the performances of models, and eightfold cross-validation was used. A mean accuracy of 98.50%. A convolutional neural network without pre-processing and with minimised layers is capable of detecting COVID-19 in a limited number of, and in imbalanced, chest X-ray images. It's a fact that other respiratory syndromes such as cough and pneumonia also shows a very common subset of symptoms as shown by Covid-19 hence there is a huge possibility that even doctors might overlook the delicacies of it and it will be very hard to differentiate between Covid and these diseases. With the growing demand of vaccine, it is even more important to detect Covid and these diseases separately. Along with the detection of Covid, this model will also be detecting these diseases and separating it from the Covid-19 chunk.

Keywords

COVID-19, convolutional neural network, X-ray, coronavirus, Cough, Pneumonia.

I. INTRODUCTION

Corona viruses are a family of hundreds of viruses that can cause fever, respiratory problems, and sometimes gastrointestinal symptoms too. The 2019 novel corona virus is one of seven members of this family known to infect humans, and the third in the past three decades to jump from animals to humans. Since emerging in China in December, this new corona virus has caused a global health emergency, sickening almost 200,000 people worldwide, and so far killing more than 9,000. As of March 19, about 10000 cases had been reported in the assumed to be 20 times this number. The basic reproduction number (BRN) is the expected number of cases directly generated by one case. A BRN greater than one indicates that the outbreak is self-sustaining, while a BRN less than one indicates that the number of new cases decreases over time and eventually the outbreak will stop. Ideally, the BRN should be reduced in order to slow down an epidemic. The BRN in the first three phases was estimated to be 3.1, 2.6, and 1.9, respectively. In the Cell Discovery article, the BRN is assumed to have decreased to 0.9 or 0.5 in phase IV, based on previous experience in SARS. According to an article in Science in 2003, the BRN of SARS decreased from 2.7 to 0.25 after the patients were isolated and the infection started being controlled.

The better we can track the virus, the better we can fight it. By analysing different parameters responsible for the outbreak of corona virus, we can take controlling measures in an accelerated way. Pneumonia can be a life-threatening illness if not diagnose properly and can result in the death of a person associated with this kind of ailment. It is in a form of severe respiratory illness caused by transmittable agents like viruses, or bacteria that affects the lungs. It can be spread through the nose or throat and affect the lungs if they are inhaled or communicated through air-borne droplets from a person coughing or sneezing. The lungs of a person are made up of small sacs or alveoli that supplies the air passage whenever a well fit person breathes. When a person is infected with pneumonia, it limits the oxygen intake and makes breathing difficult and painful due to tissue soreness caused by alveoli covered with fluids or pus. An aging person from 50 years of age and above and kids under five years of age are susceptible to pneumonia illness for they have a weaker immune system and it has taken over a million lives globally. In the Philippines, it has reported nearly 58,000 mortalities in 2016 and the 3rd top killer behind heart diseases and cancer. COVID-19 signs and indications are almost identical to the pneumonia, if not properly diagnose will lead to incorrect diagnosis now that many hospitals around the world are congested. Many of these hospitals are working 24/7 due to massive increase of infections and most of its medical personnel are also infected with the virus. Imprecise findings of pneumonia or non-COVID-19 may be labelled incorrectly as COVID-19 infected and setbacks in proper treatment are costly, the struggle and risk of being exposed to other positive patients of COVID-19.

Convolutional Neural Networks (CNNs) have shown to be tremendously valuable in feature-extraction and learning through training and for that reason, it is commonly implemented in medical researches. The application of CNN has improved the image attributes in the environment with low-light conditions, efficient endoscopy video, lung nodule detection and identification thru computed tomography images, analysis of pneumonia through X-ray image of chest, and other pulmonary related studies. Methods of deep learning covering the deep CNNs techniques on X-Ray images of chest are receiving recognition and encouraging outcomes has made it known in diverse applications.

DATASET

The dataset used here is the real dataset. The images are gathered from Github and Kaggle and the images are type Normal chest X-Ray of people who are not affected and of those people who are affected and are COVID-19 positive. Along with this X-Rays of those are also taken who are affected by cough or Pneumonia.

II. LITERATURE SURVEY

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers. In this section, we briefly review the related work on Chest X-Rays and the techniques used. There are many factors that have an impact on this model of predicting the disease. Discovering corona virus manually is a difficult task and generally not very accurate and also not very safe, hence this model is developed to help the doctors and health workers to identify the disease without getting in contact with the patient. **Boran Sekeroglu and Ilker Ozsahin** had proposed an advanced Covid virus prediction model using CNN. This system's aim was to make a model that work on three classes: covid-19/Normal, Covid-19/Pneumonia and Covid-19/Pneumonia/Normal.

This paper performed on thirty-eight experiments were performed using convolutional neural networks, 10 experiments were performed using five machine learning models, and 14 experiments were performed using the state-of-the-art pre-trained networks for transfer learning. Images and statistical data were considered separately in the experiments to evaluate the performances of models, and eightfold cross-validation was used. A mean sensitivity of 93.84%, mean specificity of 99.18%, mean accuracy of 98.50%, and mean receiver operating characteristics–area under the curve scores of 96.51% are achieved.

The paper proposed by **Pranav Rajpurkar * 1 Jeremy Irvin * 1 Kaylie Zhu 1 Brandon Yang 1 Hershel Mehta 1 Tony Duan 1 Daisy Ding 1 Aarti Bagul 1 Robyn L. Ball 2 Curtis Langlotz 3 Katie Shpanskaya 3 Matthew P. Lungren 3 Andrew Y. Ng 1** developed an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-Ray dataset, containing over 100,000 frontal view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases. They also show that a simple extension of our algorithm to detect multiple diseases outperforms previous state of the art on ChestX-ray14, the largest publicly available chest X-Ray dataset. With automation at the level of experts, we hope that this technology can improve healthcare delivery and increase access to medical imaging expertise in parts of the world where access to skilled radiologists is limited.

The paper proposed by **Sammy V. Militante, Nanette V. Dionisio and Brandon G. Sibbaluca** explains in details the delicacies of pneumonia which is again of two types: one is the viral and other is the bacterial one along with Covid-19. s. The developed CNN model was effective in extracting features from an x-ray image and forecast the occurrence or nonexistence of COVID-19, bacterial, and viral-pneumonia. Likewise, testing-data in the research was intensified through data augmentation techniques. In addition to the improvement of computer related applications in the medical division, COVID-19 and pneumonia can be efficiently found employing chest radiographs with the support of CNN and deep learning technologies. Methodologies developed in the conduct of this research in which COVID-19, bacterial, and viral-pneumonia can be forecast with greater accuracy, and in this case the study obtained 95% accuracy.

III.METHODOLOGY

• Collection of Data

Data is the heart of machine learning. Predictive models use data for training which gives somewhat accurate results. Without data we can't train the model. Machine learning involves building these models from data and uses them to predict new data. Deep CNNs indeed achieve better performance using a large dataset compared to a smaller one. Granting there are a significant number of infected COVID-19 patients globally, but the number of publicly available chest X-ray images on- line are insignificant and dispersed. Hence, a reasonably large dataset of COVID-19 infected chest X-ray images has been collected although normal and pneumonia images are promptly accessible publicly and applied in this study. The dataset used here is taken from Kaggle and GitHub and some other platforms.

The covid chestxray dataset is taken from this GitHub repository

https://github.com/ieee8023/covid-chestxray-dataset

This dataset consists of X-ray and CT scan images of patients infected to COVID-19, SARS, Streptococcus, ARDS, Pneumocystis, and other types of pneumonia from different patients.

The second dataset was taken from

https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

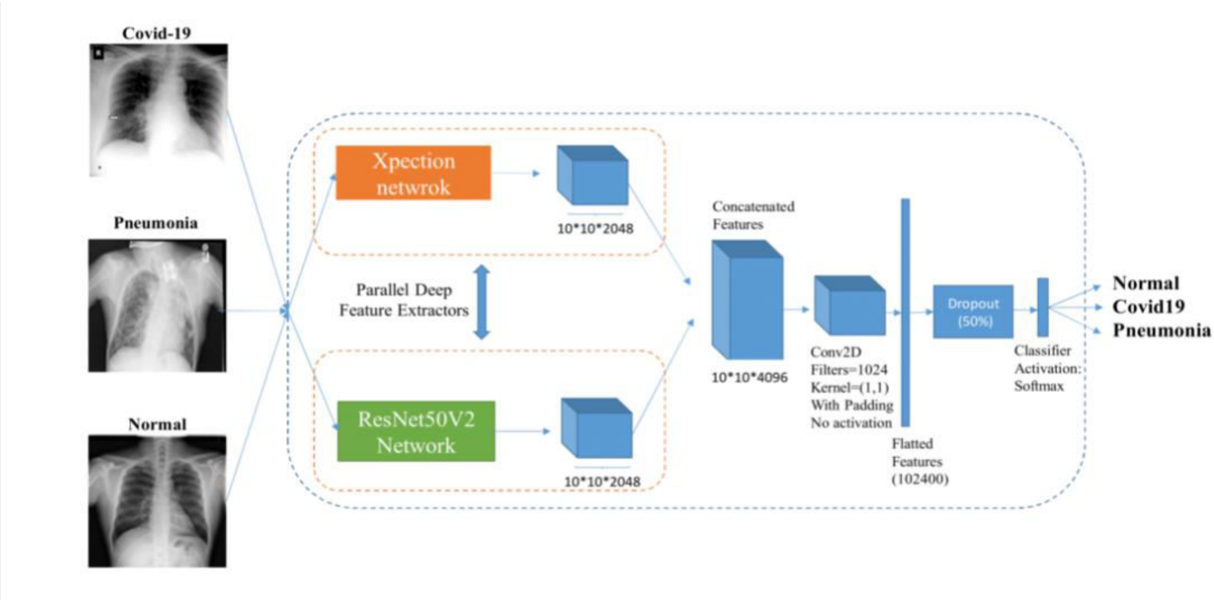
which contains 6012 cases with pneumonia and 8851 normal cases.

• Building and Training the model

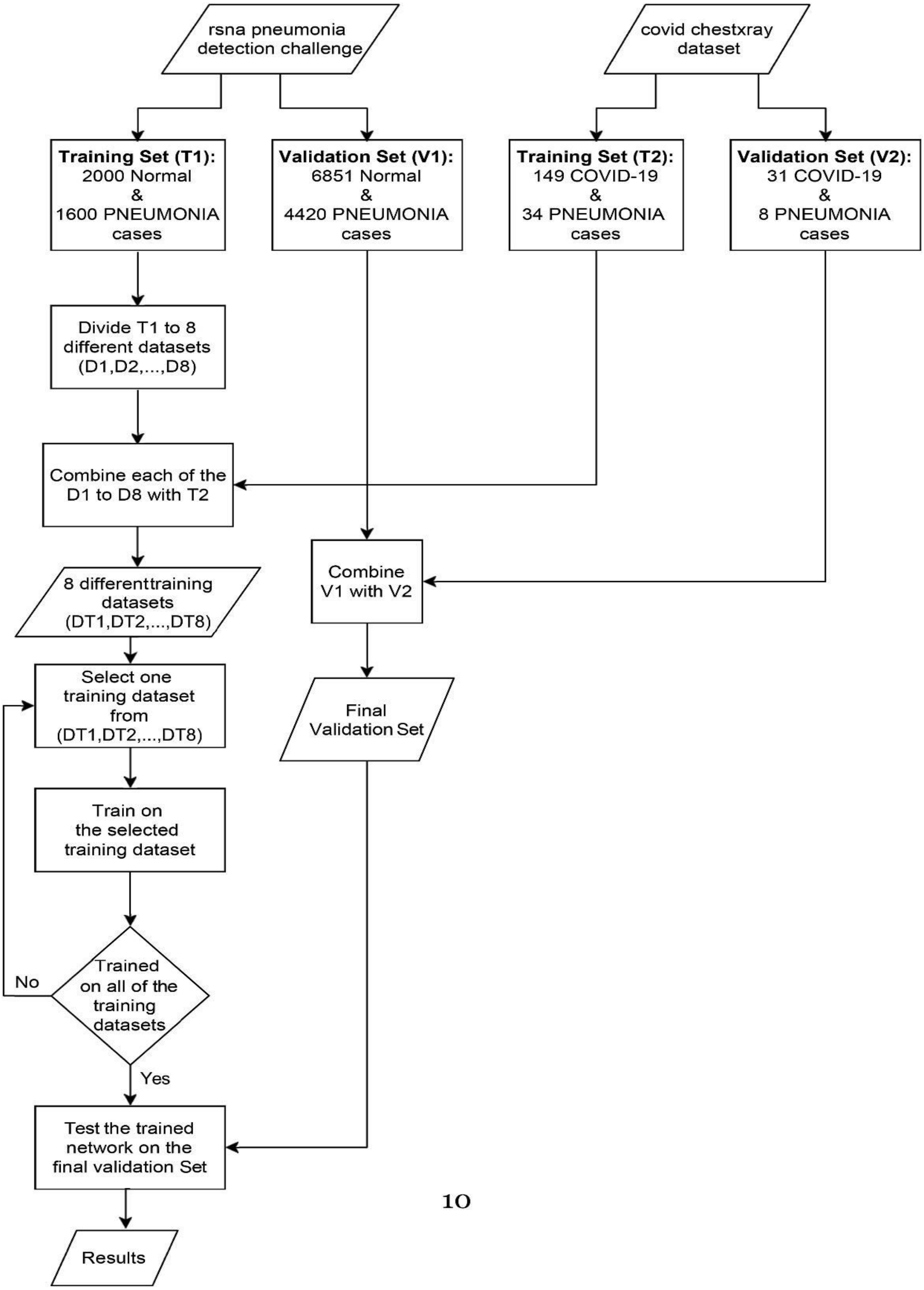
The pre-processed input images of our dataset are 300 × 300 pixels. Xception generates a 10 × 10 × 2048 feature map on its last feature extractor layer from the input image, and ResNet50V2 also produces the same size of feature map on its final layer. As both networks generate the same size of feature maps, I concatenated their features so that by using both of the inception-based layers and residual-based layers, the quality of the generated semantic features would be enhanced.

A concatenated neural network is designed by concatenating the extracted features of Xception and ResNet50V2 and then connecting the concatenated features to a convolutional layer that is connected to the classifier. The kernel size of the convolutional layer that was added after the concatenated features was 1 × 1 with 1024 filters and no activation function. This layer was added to extract a more valuable semantic feature out of the features of a spatial point between all channels, with each channel being a feature map. This convolutional layer helps the network learn better from the concatenated features extracted from Xception and ResNet50V2.

The architecture of the proposed model:

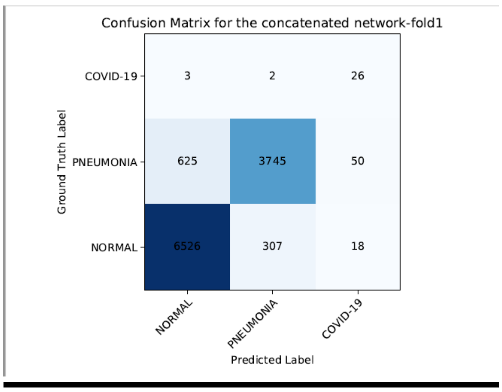


The flowchart of the proposed method:

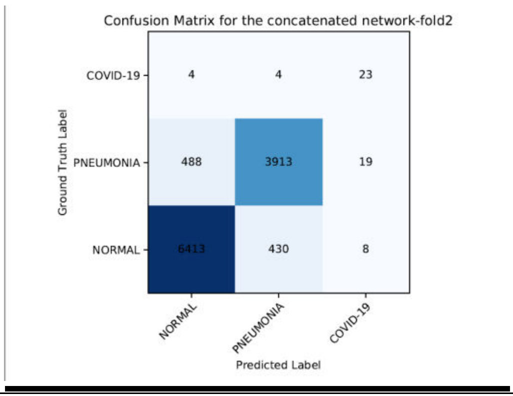


10

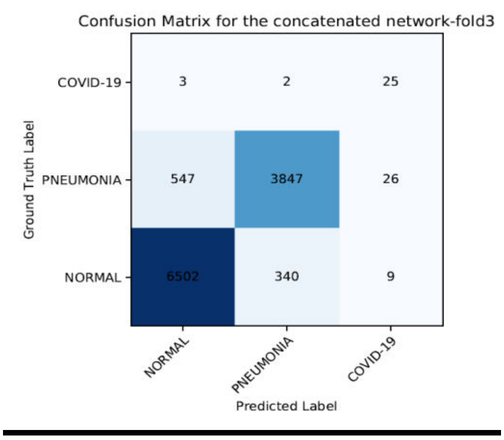
IV.RESULTS



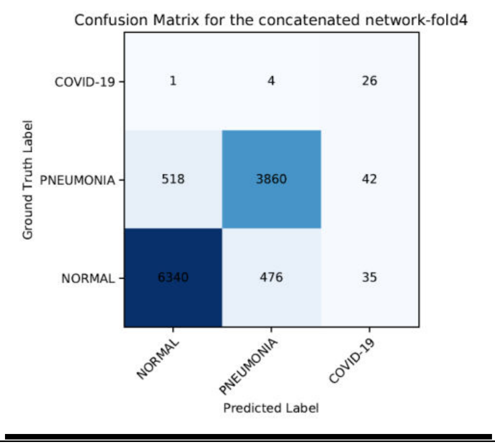
(Confusion Matrix for the concatenated network-fold1)



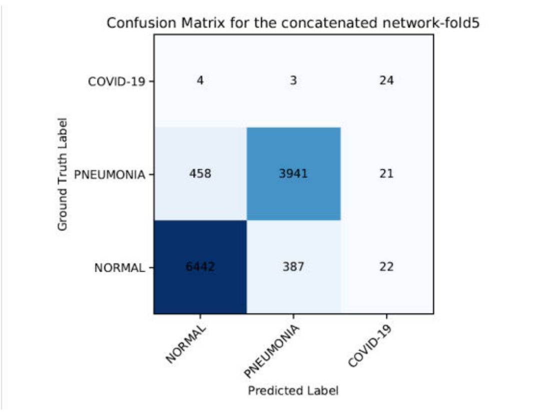
(Confusion Matrix for the concatenated network-fold2)



(Confusion Matrix for the concatenated network-fold3)



(Confusion Matrix for the concatenated network-fold4)



(Confusion Matrix for the concatenated network-fold5)

V.CONCLUSION

Detection of COVID-19 from chest X-ray images is of vital importance for both doctors and patients to decrease the diagnostic time and reduce financial costs. Artificial intelligence and deep learning are capable of recognizing images for the tasks taught. In this study, several experiments were performed for the high-accuracy detection of COVID-19 in chest X-ray images. Various groups—COVID-19/Normal and COVID-19/Positive cough and pneumonia were considered for the classification. Different image dimensions, different network architectures, and machine learning models were implemented and evaluated using images and statistical data. When the number of images in the database and the detection time of COVID-19 are considered, it can be suggested that the considered architectures reduce the computational cost with high performance. The results showed that the convolutional neural network with minimised convolutional and fully connected layers is capable of detecting COVID-19 images within the classes, COVID-19/Negative and COVID-19/ Positive and along with it cough and pneumonia. Therefore, the use of AI-based automated high-accuracy technologies may provide valuable assistance to doctors in diagnosing COVID-19, another lung related disease. Further studies, based on the results obtained in this study, would provide more information about the use of CNN architectures with COVID-19 chest X-ray images and improve on the results of this study. In addition to the improvement of computer- related applications in the medical division, COVID-19 and pneumonia can be efficiently found employing chest radiographs with the support of CNN and deep learning technologies.

VI. REFERENCES

[1] Anuj V. Kumar1, Anshuman Kumar2, Satish S. Tiwari 3, Sneha G. Gobade4, Prof. Amita Suke5, "HOUSE COST PREDICTION USING DATA SCIENCE AND MACHINE LEARNING" Volume: 07 Issue: 01 | Jan 2020

[2] PropTech for Proactive Pricing of Houses in Classified Advertisements in the Indian Real Estate Market Sayan Putatunda Member, IEEE

[3] F. Wu, S. Zhao and B. Yu, "A new coronavirus associated with human respiratory disease in China," Nature, vol. 579, pp. 265-269, 2020.

[4] Wei-jie Guan, Ph.D., Zheng-yi Ni, M.D., Yu Hu, M.D., Wen-hua Liang, Ph.D., Chun-quan Ou, Ph.D., Jian-xing He, M.D., Lei Liu, M.D., Hong Shan, M.D., Chun-liang Lei, M.D., David S.C. Hui, M.D., Bin Du, M.D., Lan-juan Li, M.D., et al., for the China Medical Treatment Expert Group for Covid-19. N Engl J Med 2020; 382:1708-1720 DOI: 10.1056/NEJMoa2002032

[5] Zhou, P., Yang, XL., Wang, XG. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579,

270–273 (2020).

[6] Cleverley Joanne, Piper James, Jones Melvyn M. The role of chest radiography in confirming covid-19 pneumonia BMJ 2020; 370 :m2426

[7] Chest X-ray images (Normal and Pneumonia): https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[8] Marcin Kociołek, Michał Strzelecki, Rafał Obuchowicz, Does image normalization and intensity resolution impact texture classification?, Computerized Medical Imaging and Graphics, Volume 81, 2020, 101716, ISSN 0895-6111

[9] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).

[10] Perez, G., Arbelaez, P. Automated lung cancer diagnosis using threedimensional convolutional neural networks. Med Biol Eng Comput 58, 1803–1815 (2020).

[11] For dataset: http://cb.lk/covid_19

[12] https://github.com/ieee8023/covid-chestxray-dataset

[13] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

[14] F. Wu, S. Zhao and B. Yu, "A new coronavirus associated with human respiratory disease in China," Nature, vol. **579**, pp. 265–269, 2020.

[15] Wei-jie Guan, Ph.D.., Zheng-yi Ni, M.D.., Yu Hu, M.D.., Wen-hua Liang, Ph.D.., Chun-quan Ou, Ph.D.., Jian-xing He, M.D.., Lei Liu, M.D.., Hong Shan, M.D.., Chun-liang Lei, M.D.., David S.C. Hui, M.D.., Bin Du, M.D., Lan-juan Li, M.D.., et al., for the China Medical Treatment Expert Group for Covid-19. N Engl J Med 2020; **382**:1708–1720 DOI: 10.1056/NEJMoa2002032

[16].Rajendran, D.K., Rajagopal, V., Alagumanian, S. et al. Systematic literature review on novel corona virus SARS-CoV-2: a threat to human era. VirusDis. **31**, 161–173 (2020).

[17] Su, S., Wong, G., Shi, W., Liu, J., Lai, A., Zhou, J., Liu, W., Bi, Y., & Gao, G. F. (2016). Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends in microbiology, **24**(6), 490–502.

[18] Zhou, P., Yang, XL., Wang, XG. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **579**, 270–273 (2020).