Open access • Posted Content • DOI:10.20944/PREPRINTS202003.0300.V1

Detection of Coronavirus Disease (COVID-19) Based on Deep Features

Prabira Kumar Sethy, Santi Kumari Behera

Published on: 19 Mar 2020

Topics: Coronavirus

Detection of coronavirus Disease (COVID-19) based on Deep Features

Prabira Kumar Sethy *[a], Santi Kumari Behera [b]

Department of Electronics, Sambalpur University, Odisha, India

Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Odisha, India

Email-prabirsethy.05@gmail.com

Abstract- The detection of coronavirus (COVID-19) is now a critical task for the medical practitioner. The coronavirus spread so quickly between people and approaches 100,000 people worldwide. In this consequence, it is very much essential to identify the infected people so that prevention of spread can be taken. In this paper, the deep learning based methodology is suggested for detection of coronavirus infected patient using X-ray images. The support vector machine classifies the corona affected X-ray images from others using the deep feature. The methodology is beneficial for the medical practitioner for diagnosis of coronavirus infected patient. The suggested classification model, i.e. resnet50 plus SVM achieved accuracy, FPR, F1 score, MCC and Kappa are 95.38%,95.52%, 91.41% and 90.76% respectively for detecting COVID-19 (ignoring SARS, MERS and ARDS). The classification model ResNet50 plus SVM is superior compared to other classification models. The result is based on the data available in the repository of GitHub, Kaggle and Open-i as per their validated X-ray images.

Keyword- coronavirus, COVID-19, diagnosis, deep features, SVM.

1. Introduction

Nowadays, "coronavirus" is frequently introduced with the word "novel," as it is a new strain in the family of viruses, we have all observed previously. As indicated by the WHO, coronaviruses belong to a large family range from the common cold to dangerous diseases [1]. These diseases can infect both humans and animals. The strain that started spreading in Wuhan, the capital of China's Hubei region, is identified from two different coronaviruses, i.e. severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS). Symptoms of coronavirus contamination go in seriousness from respiratory complication like pneumonia, kidney disorder and development of liquid in the lungs.

On February 11, 2020, the WHO Director-General, Dr. Tedros Adhanom Ghebreyesus, reported that the infection brought about by this new CoV was a "COVID-19," which is the abbreviation of "coronavirus disease 2019". In the last two decades, two coronavirus epidemics are observed, i.e. SARS-CoV and MERS-CoV. The first one started in china, spread to twenty-four countries and reported 8000 cases & 800 of deaths. The second one started from Saudi Arabia, reported 2500 cases and 800 deaths. The detail of coronavirus is depicted in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

Table 1. Detail of coronavirus

CoV	Year	Origin	Mortality rate
SARS	2002	Guangdong province, China	10%
MERS	2013	Saudi Arabia	34%
COVID-19	2019	Wuhan, China	3.4%

Among the causing pathogens for respiratory diseases, CoV is become the dangerous one because of its serial interval (5 to 7.5) and reproductive rate (2 to 3) [2]. The CoV belongs to single-stranded RNA viruses (+ssRNA) family mostly observed in animals [3,4]. The analysis carried out till date, the viruses have no species barrier and can cause severe diseases like MERS and SARS. The coronavirus infection can provoke SARS that is severe enough to be called Acute respiratory distress syndrome (ARDS). In general, estimates suggest that 2% of the population are healthy carriers of a CoV and that these viruses are responsible for about 5% to 10% of acute respiratory infections [5]. COVID-19 spreads more easily than SARS and have symptoms like other coronaviruses. Figure 1 shows the distribution of COVID-19 cases worldwide, as of 17 March 2020 [6]. Figure 2 shows the distribution of COVID-19 cases by continent (except China), as of 17 March 2020 (according to the applied case definition and testing strategies in the affected countries) [6].

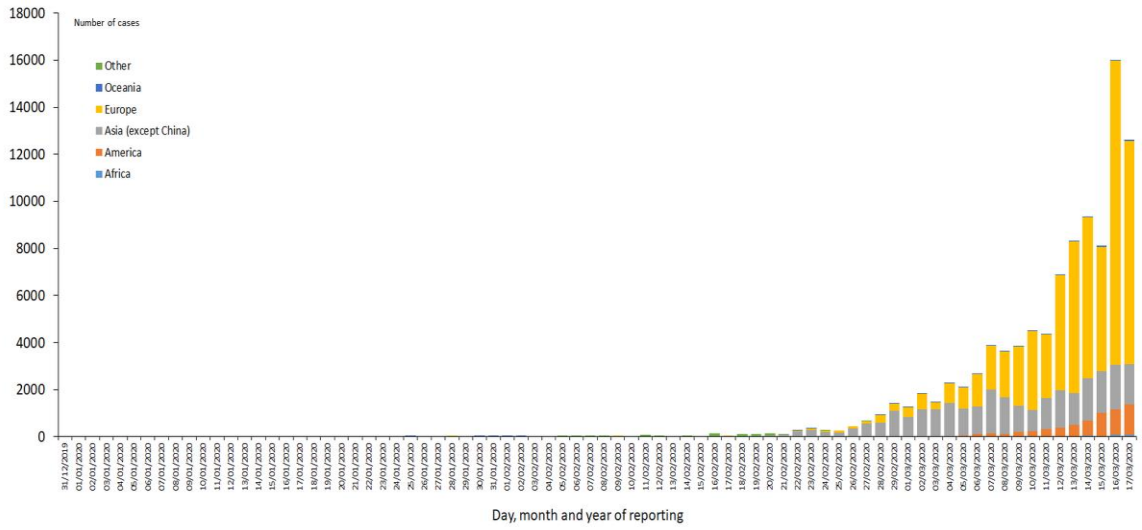

Figure 1. Distribution of COVID-19 cases worldwide, as of 17 March 2020.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

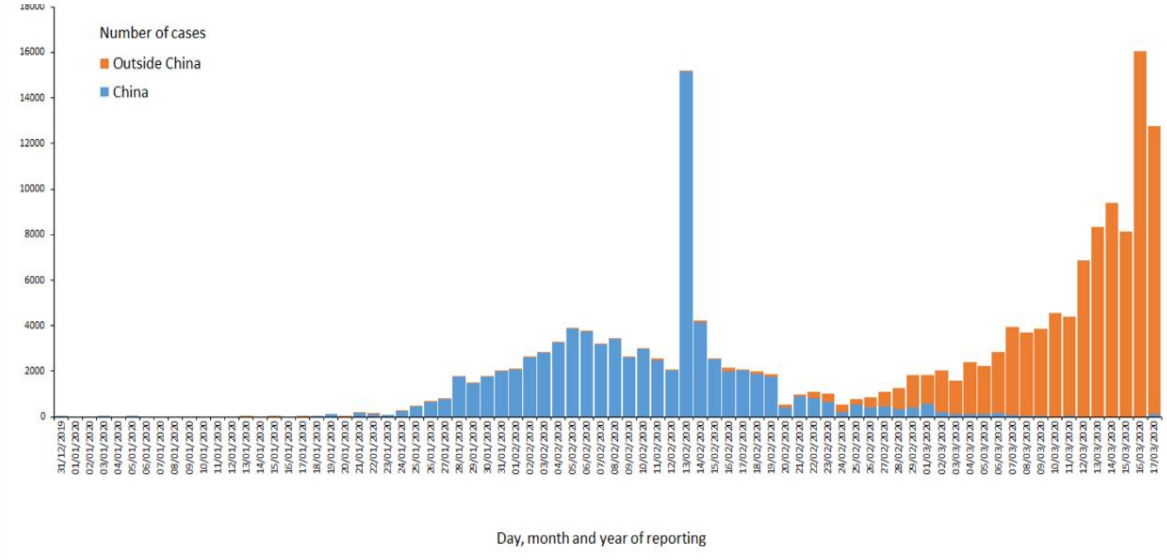

Figure 2. Distribution of COVID-19 cases by continent (except China), as of 17 March 2020 (according to the applied case definition and testing strategies in the affected countries).

The test of COVID-19 is currently a difficult task because of unavailability of diagnosis system everywhere, which is causing panic. Because of the limited availability of COVID-19 testing kits, we need to rely on other diagnosis measures. Since COVID-19 attacks the epithelial cells that line our respiratory tract, we can use X-rays to analyse the health of a patient's lungs. The medical practitioner frequently uses X-ray images to diagnose pneumonia, lung inflammation, abscesses, and/or enlarged lymph nodes. And almost in all hospitals have X-ray imaging machines, it could be possible to use X-rays to test for COVID-19 without the dedicated test kits. Again, a drawback is that X-ray analysis requires a radiology expert and takes significant time, which is precious when people are sick around the world. Therefore, developing an automated analysis system is necessary to save medical professionals valuable time. The chest X-ray images of COVID-19^{+} are shown in Figure 1.

In the current situation of the rapid spread of COVID-19 many kinds of research have been going on [7-15]. The deep learning is one of the recent techniques applicable in the field of medicine for diagnosis purpose [16-28].

In this paper, a system based on deep CNNs is developed for the identification of COVID-19 as a classification task. In this study, we prepared two sets of datasets. The first dataset contains 25 number of COVID-19^{+} and 25 number of COVID-19^{-} X-ray images. The COVID-19^{+} X-ray images are collected from the GitHub repository shared by Dr. Joseph Cohen, a postdoctoral fellow at the University of Montreal [29]. The COVID-19^{-} are the X-ray images of pneumonia collected from Kaggle repository [30]. The COVID-19^{+} excludes the MERS, SARS, and ARDS. The second dataset contains 133 X-ray images of COVID-19^{+}, including MERS, SARS, and ARDS. In addition, 133 chest X-ray images are collected from the Open-i repository [31] as COVID-19^{-}. The two datasets are examined separately in the proposed models. We use this dataset for deep feature extraction based on deep

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

learning architectures such as AlexNet, VGG16, VGG19, GoogleNet, ResNet18, ResNet50, ResNet101, InceptionV3, InceptionResNetV2, DenseNet201 and XceptionNet. The deep features obtained from these deep models are classified by SVM. Again, the transfer learning approach is applied for the identification of rice diseases in the aforementioned deep CNN models. Finally, we evaluate the performance results by using deep feature extraction methods. The detail about the dataset is in Table 2.

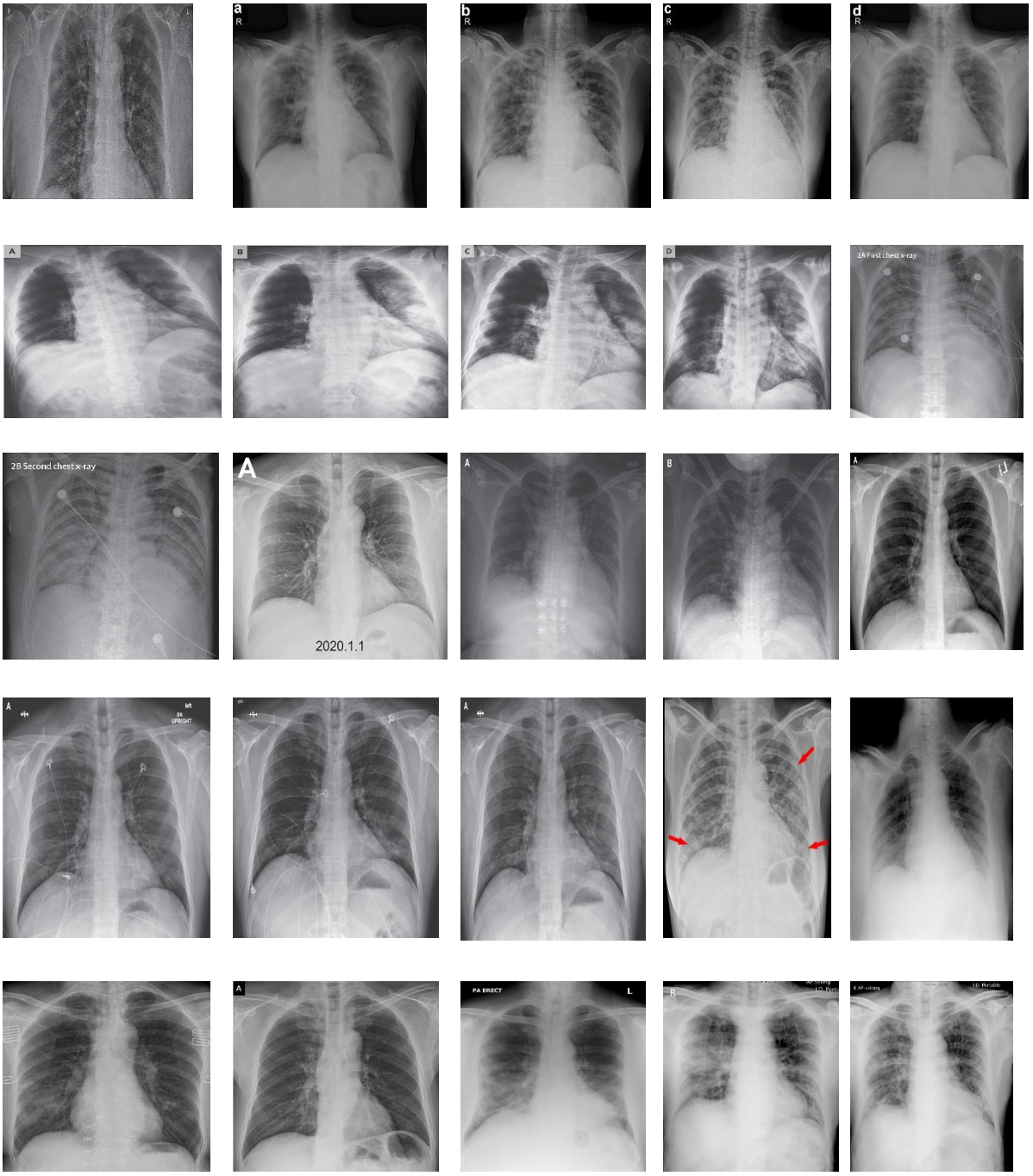

Figure 1. X-ray images of COVID-19^{+} ignoring MERS, SARS, and ARDS.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

Table 2 Detail of Dataset.

Samples		Number	Repository
	COVID-19+ without MERS, SARS, ARDS	25	GitHub (Dr. Joseph Cohen)
	COVID-19-	25	Kaggle (X-ray images of Pneumonia)

2. Methodology

Deep feature extraction is based on the extraction of features acquired from a pre-trained CNN [32]. The deep features are extracted from fully connected layer and feed to the classifier for training purpose. The deep features obtained from each CNN networks are used by SVM classifier. After that, the classification is performed, and the performance of all classification models are measured. The rice leaf disease identification model based on deep features by SVM classifier is shown in Figure 2.

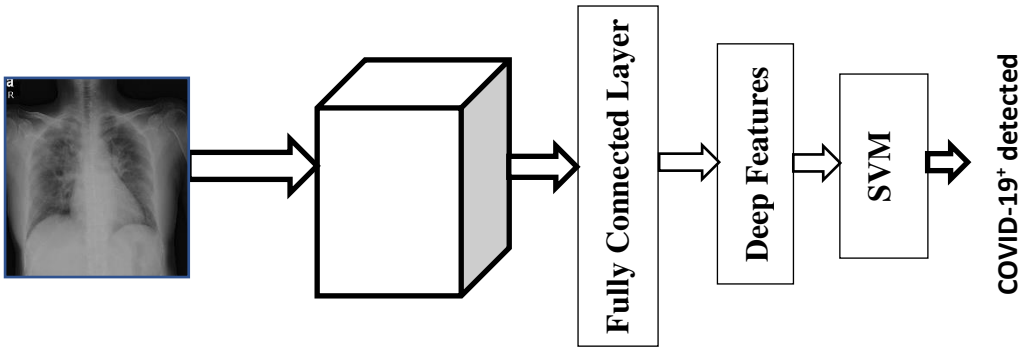

Figure 2. Detection Corona Virus by SVM based on Deep Feature using X-ray images.

The deep features of CNN models are extracted from a particular layer and feature vector is obtained. The features are fed to the SVM classifier for rice disease identification purpose. The feature layer and feature vector are detailed in Table 3.

Table 3. Details of feature layer and feature vector of CNN models.

CNN models	Feature Layer	Feature Vector	CNN models	Feature Layer	Feature Vector
AlexNet	fc6	4096	Xception	predictions	1000
			Resnet18	Fc1000	1000
			Resnet50	Fc1000	1000
Vgg16	fc6	4096	Resnet101	Fc1000	1000
			Inceptionv3	predictions	1000
			Inceptionresnetv2	predictions	1000
Vgg19	fc6	4096	GoogleNet	loss3-classifier	1000
			Densenet201	Fc1000	1000

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

3. Results and Discussion

In this study, we examined the performance of classification models for identification COVID-19+ based on eleven number of CNN models. The experimental studies were implemented using the MATLAB 2019a deep learning toolbox. All applications were run on a laptop, i.e. Acer Predator Helios 300 Core i5 8th Gen - (8 GB/1 TB HDD/128 GB SSD/Windows 10 Home/4 GB Graphics) and equipped with NVIDIA GeForce GTX 1050Ti. The measurement of performance of each classifier is measured in terms of Accuracy, Sensitivity, Specificity, False positive rate (FPR), F1 Score, MCC and Kappa. In addition, this experimentation used One-Vs-all approach and linear SVM as the SVM classifier parameter. The results reported in Table 4 and Table 5 are based on the average value of 100 independent simulations. The training, validation and testing ration is 60:20:20 and adapted randomized selection for training, validation and testing in each execution. The results reported in Table 4 and Table 5 are the cases of coronavirus excluding SARS, MERS and ARDS i.e. only for COVID-19.

Note: The result is based on the data available in the repository of GitHub, Kaggle and Open-i as per their validated X-ray images. The author is not responsible if the data uploaded in the repository are incorrect.

Table 4 Statistical analysis of different classification models based on SVM using deep features of various CNN models. (Accuracy, Sensitivity and Specificity)

Classification Models	Accuracy	Sensitivity	Specificity
AlexNet	.9332352941^{d}	.9341176471^{b,c}	.9323529412^{c}
DenseNet201	.9388235294^{d}	.9435294118^{c,d}	.9341176471^{c}
GoogleNet	.9144117647^{b,c}	.8982352941^{a}	.9305882353^{b,c}
Inceptionv3	.9108823529^{b}	.9111764706^{a,b}	.9105882353^{b,c}
ResNet18	.9108823529^{b}	.9111764706^{a,b}	.9105882353^{b,c}
ResNet50	**.9538235294^{e}**	.9729411765^{d,e}	**.9347058824^{c}**
ResNet101	.8926470588^{a}	.9123529412^{a,b}	.8729411765^{a}
VGG16	.9276470588^{c,d}	**.9747058824^{e}**	.8805882353^{a}
VGG19	.9291176471^{d}	.9511764706^{c,d,e}	.9070588235^{b}
XceptionNet	.9391176471^{d}	.9476470588^{c,d,e}	.9305882353^{b,c}
Inceptionresnetv2	.9332352941^{d}	.9529411765^{c,d,e}	.9135294118^{b,c}

Means within a column the same letter(s) are not statistically significant (p=0.05) according to Duncan's multiple range test (SPSS Version 26).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2020 doi:10.20944/preprints202003.0300.v1

Table 5 Statistical analysis of different classification models based on SVM using deep features of various CNN models (FPR, F1 Score, MCC and Kappa).

Classification Models	FPR	F1 Score	MCC	Kappa
AlexNet	.0676470588[a]	.9278974778[c]	.8779026733[c]	.8664705882[d]
DenseNet201	.0658823529[a]	.9376506154[c]	.8839241390[c]	.8776470588[d]
GoogleNet	.0694117647[a,b]	.9111546673[b]	.8381731432[b]	.8288235294[b,c]
Inceptionv3	.0894117647[a,b]	.9093669453[c]	.8319417112[b]	.8217647059[b]
ResNet18	.0894117647[a,b]	.9093669453[b]	.8319417112[b]	.8217647059[b]
ResNet50	**.0652941176[a]**	**.9552269690[d]**	**.9141314397[d]**	**.9076470588[e]**
ResNet101	.1270588235[b]	.8901809196[a]	.7983680368[a]	.7852941176[a]
VGG16	.1194117647[b]	.9313784515[c]	.8646114190[c]	.8552941176[c,d]
VGG19	.0929411765[b]	.9294454314[c]	.8651893172[c]	.8582352941[d]
XceptionNet	.0694117647[a,b]	.9385235325[c]	.8822976830[c]	.8782352941[d]
Inceptionresnetv2	.0864705882[a,b]	.9343053714[c]	.8707484387[c]	.8664705882[d]

Means within a column the same letter(s) are not statistically significant (p=0.05) according to Duncan's multiple range test (SPSS Version 26).

It is observed from Table 4, the accuracy of ResNet50 plus SVM is superior to other classification models (having different superscript in the column, i.e. letter 'e'). It is observed from Table 5, in terms of F1score, MCC and Kappa, the ResNet50 plus SVM classification model is statistically superior to the other classification models. Hence, resnet50 and SVM result better classification for detection of COVID-19^{+} with accuracy, FPR, F1 score, MCC and Kappa are 95.38%,95.52%, 91.41% and 90.76% respectively.

4. Conclusion

The content of the manuscript about the coronavirus is based on the data available in WHO, European Centre for Disease Prevention and Control An agency of the European Union and other official websites worldwide. The chest X-ray images of are used for simulation purposes are collected from GitHub, Kaggle and Open-I repository. For detection of coronavirus using X-ray images based on deep feature and SVM. For this, we extracted the deep feature of nine pretrained CNN model and fed to SVM classifier individually. To choose the best classification model, statistical analysis is carried out. The classification model, i.e. ResNet50 plus SVM is statistically superior compare to other eight models. The proposed classification model for detection of COVID-19 is achieved 95.38% of accuracy.

References

[1] www.who.int [Accessed:17th March 2020]

[2] Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. (2020). Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. doi:10.1016/j.ijid.2020.02.060

[3] Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. Nat. Rev. Microbiol. 2009 Jun;7(6):439-50.

[4] Chan JF, To KK, Tse H, Jin DY, Yuen KY. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol. 2013 Oct;21(10):544-55.

[5] Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J. Med. Virol. 2020 Apr;92(4):418-423.

[6] European Centre for Disease Prevention and Control An agency of the European Union. Link: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

[7] Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv. 2020

[8] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama. 2020

[9] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020.

[10] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. The New England journal of medicine. 2020.

[11] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020.

[12] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2020; 25(3).

[13] Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, et al. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clinical chemistry. 2020.

[14] Zhang N, Wang L, Deng X, Liang R, Su M, He C, et al. Recent advances in the detection of respiratory virus infection in humans. J Med Virol. 2020.

[15] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology. 2020: 200230.

[16] Gomez P, Semmler M, Schutzenberger A, Bohr C, Dollinger M. Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. Med Biol Eng Comput. 2019; 57(7): 1451-63.

[17] Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. Radiology. 2019; 292(2): 365-73.

[18] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 2018; 172(5): 1122-31.

[19] Negassi M, Suarez-Ibarrola R, Hein S, Miernik A, Reiterer A. Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects. World J Urol. 2020.

[20] Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng. 2018; 2(10): 741-8.

[21] Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH. Radiographic and CT Features of Viral Pneumonia. Radiographics. 2018; 38(3): 719-39.

[22] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in Biology and Medicine, 100, 270–278. doi:10.1016/j.compbiomed.2017.09.017

[23] Yıldırım, Ö., Pławiak, P., Tan, R.-S., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine. doi:10.1016/j.compbiomed.2018.09.009

[24] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. Computers in Biology and Medicine, 89, 389–396. doi:10.1016/j.compbiomed.2017.08.022

[25] Oh, S. L., Ng, E. Y. K., Tan, R. S., & Acharya, U. R. (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in Biology and Medicine. doi:10.1016/j.compbiomed.2018.06.002

[26] Burlina, P., Pacheco, K. D., Joshi, N., Freund, D. E., & Bressler, N. M. (2017). Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. Computers in Biology and Medicine, 82, 80–86. doi:10.1016/j.compbiomed.2017.01.018

[27] Xia, Y., Wulan, N., Wang, K., & Zhang, H. (2018). Detecting atrial fibrillation by deep convolutional neural networks. Computers in Biology and Medicine, 93, 84–92. doi:10.1016/j.compbiomed.2017.12.007

[28] Wahab, N., Khan, A., & Lee, Y. S. (2017). Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. Computers in Biology and Medicine, 85, 86–97. doi:10.1016/j.compbiomed.2017.04.012

[29] www.github.com [Accessed: 16th March 2020]

[30] www.kaggle.com [Accessed:16th March 2020]

[31] https://openi.nlm.nih.gov [Accessed:16th March 2020]

[32] Lopes UK, Valiati JF. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Computers in biology and medicine. 2017 Oct 1;89: 135-43.