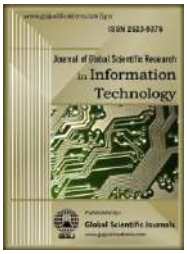
Journal of Global Scientific Research 7 (4) 2023/ 3056-3065

Contents lists available at www.gsjpublications.com

Journal of Global Scientific Research in Information Technology

journal homepage: www.gsjpublications.com/jgsr



A New Data Classification Approach Based on Deep Learning Techniques

Sabah B. Rashed, Yasemin GÜLTEPE

Computer Engineering Department, Kastamonu University, Kastamonu, Turkey

ARTICLE INFO

Received: 22 Feb 2023,
Revised: 23 Feb 2023,
Accepted: 5 Mar 2023,
Online: 9 Apr 2023

Keywords:
Data classification, Deep learning, Covid-19, x-ray

ABSTRACT

In this study, we developed a new model based on Convolutional Neural Network (CNN) and the power of signal function for Covid-19 detection. The presented method consists of three parts: CNN, power of the signal, and classifiers. The aim of applying CNN is to extract high-level and sensitive features from input x-ray images. The CNN pre-trained model Alexnet is used in this section as a features extractor and the extracted features are wired to the power of signal function that is used to calculate the power of each period and reduce the size of input features. Then, the extracted features by the power of the signal are wired directly to the classifiers. The obtained results have 95.5% accuracy which is remarkable when compared with several states of art studies presented in this field.

1. Introduction

Data classification is the process of sorting and classifying data into different types, forms, or other distinct categories. Data classification enables the separation and classification of data according to the requirements of the data set for various business or personal purposes. It is basically the process of managing data [1, 2]. Deep learning is a section or part of machine learning that is part of the larger science called Artificial Intelligence.

It is a technique invented by man in order to try to imitate the way the human mind works. Deep learning tries to simulate the human mind in all its capabilities, which include; Seeing, understanding speaking, listening, and other capabilities are powerful and unrivaled by our human mind. Not only that, but that scientists have studied the human brain and how it works in order to design algorithms and programs capable of simulating it, and for this reason, we find that these algorithms are inspired by human medical and neurological studies and try as much as possible to imitate them, but in computational rather than biological ways. Neurons or Neural Networks have been replaced by a computer to become a perceptron or Artificial neural network, of which we now have many types such as: Convolutional Neural Network, or Recurrent Neural Network [3, 4]. X-rays are a type of electromagnetic radiation that has a wavelength in the range from 10 to 0.01 nanometers, corresponding to frequencies in the range 30 petahertz to 30 exahertz (30 × 1015 Hz to 30 × 1018 Hz) and energy in the range 120 volts to 120 KV The wavelength of X-rays is shorter than that of ultraviolet rays. In many

Corresponding author:
E-mail address: sabahrashed1984@gmail.com
doi: 10.5281/jgsr.2023.7809370

languages, it is called Röntgen rays after their first discoverer, Wilhelm Röntgen, who called them X-rays, which means unknown rays. The discovery of X-rays opened the way for more accurate examination of the internal organs and viscera involved in the human body. Then the German discoverer Roentgen deserved to win the prestigious Nobel Prize in Physics in 1901 and it was the first in this branch since the award was established at all. [5, 6]. The New Coronavirus Disease (COVID-19) is a virus that was first identified on January 13, 2020 as a result of research conducted in a group of patients who developed respiratory symptoms (fever, cough, shortness of breath) in Wuhan Province in late December. The outbreak was initially detected in those in the seafood and animal market in this region [7, 8]. Later, it spread from person to person and spread to other cities in Hubei province, especially Wuhan, and other provinces of the People's Republic of China and other world countries. Coronaviruses are a large family of viruses that can cause disease in animals or humans. In humans, several coronaviruses are known to cause respiratory infections, ranging from the common cold to more severe illnesses such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) [8]. The New Coronavirus Disease is caused by the SAR-CoV-2 virus. Although asymptomatic cases have been reported, their rate is unknown. The most common symptoms are fever, cough and shortness of breath. In severe cases, pneumonia, severe respiratory failure, kidney failure and death may develop. It is transmitted by the inhalation of droplets scattered by the sick individuals with their coughing and sneezing. The virus can also be taken by taking the hands of the patients to the face, eyes, nose or mouth without washing their hands after touching surfaces contaminated with respiratory particles. It is risky to touch the eyes, nose or mouth with dirty hands [9, 10].

Related work

Since the outbreak of the pandemic, automated screening for Covid-19 disease detection has become a top priority for the research community. The development of an automated system for classifying CT scans of the lung remains challenging due to the complexity of diagnosing infectious and inflammatory lung diseases on a visual inspection. Although visual inspection is an acceptable standard, it tends to be subject to errors from the large number of patients that need to be diagnosed. In this regard, researchers have proposed many studies to automatically identify the unique features of Covid-19 by various methods. Kang et al. have proposed a multi-image representation learning technique that can automatically diagnose Covid-19 [11]. They applied their proposed model to 2522 CT scan images to validate. With the method they applied, they reached 95.5%, 96.6% and 93.2% accuracy, sensitivity and specificity, respectively. Li et al. have proposed a deep learning automated framework, COVNet, to accurately identify Covid-19 with chest CTs [12]. While creating their models, they used chest CT consisting of 4356 images. With this model, a sensitivity of 87% and an Area Under the Curve (AUC) value of 0.95 were obtained in detecting Covid-19 patients from other pneumonia patients. Xu et al. have designed a deep learning model called ResNet for early detection of Covid-19 [13]. A total of 618 pulmonary CT samples were used while creating the model. In this study, a final accuracy of 86.7% was achieved that distinguishes Covid-19 from influenza-A pneumonia and healthy cases. Ardakani et al. used ten convolutional neural networks, namely VGG-16, VGG-19, AlexNet, GoogleNet, SqueezeNet, ResNet-18, ResNet-50, ResNet101, MobileNet-V2 and Xception, to differentiate Covid-19. They compared it with other pneumonias (ie non-Covid-19) using

1020 CT images [14]. In this regard, they observed that ResNet 101 and Xception achieved the highest AUC of 0.994 and recommended Resnet 101 to characterize and detect Covid-19 patients. In another study, Bai X. et al. introduced EfficientNet, a deep neural network architecture, and applied CT frames obtained from 1186 patients to this architecture [15]. While distinguishing between Covid-19 and non-Covid-19 in the produced system, 96% accuracy, 95% sensitivity and 96% specificity were obtained in the released system. Shi et al. have applied Random Forest (RO), a machine learning algorithm, to screen for Covid19 [16]. presented for this study. They used CT images of 2685 patients to evaluate their model. In the model, after evaluating the 5-fold cross-validation technique, the model reached accuracy, sensitivity

and specificity of 87.9%, 90.7% and 83.3%, respectively. In another study, Özkaya et al. produced 3000 patch images from 150 CT images and applied further sequencing and fusion techniques on these images [17]. A Support Vector Machine (SVM) was used for classification, and before that a pre-trained ESA model was used as part of the transfer learning method. In the presented procedure, 98.27% accuracy, 97.63% precision and 97.6% sensitivity were obtained. Alom et al., [18] presented the transfer-learning, RegionalConvolutional Neural Network (B-ESA), an efficient deep learning approach to detect Covid19. Researchers have previously used the NABLA-N network for infected area segmentation to improve the outcome of classification. In order to evaluate the proposed method, both Xray and CT images were studied. They obtained 84.67% and 98.78% accuracy values from X-ray and CT images, respectively. Current classification models show some limitations in terms of feature extraction complexity. Various feature extraction algorithms play an important role in capturing significant changes in the spatial distribution of image pixels. Recently, fractional analysis and its applications have been used in different application areas [19]. In this study, existing models for lung CT scan classification for image classification tasks, considered one of the contributions of the study, rely solely on deep learning for feature extraction. Therefore, combining clinical findings with deep learning features will positively affect the course of the disease history by further improving the classification performance between − Covid-19 and healthy cases. The motivation of this study is to propose an efficient classification of healthy lungs with Covid-19 in CT scans using deep learning.

2. Methodology

In this section, the proposed model presented and explained. In the first stage, the CNN applied to extract important features and reduce the size of x-ray images from 227*227*3 to 4096*1 for each image. The CNN play important role in reducing the execution time which reduced the features size for each sample from 154,587 to 4096 features that is the output of Alexnet [15]. Then the 4096 features wired to the power of signal function which can be calculated for 2 period or 4 period. The power calculated according to the equation (1):

$p = \lim^{1}$ $^{-}(t)\backslash^{2}$ (1)

T $\backslash x_{T}$

X (t) represented the vector of features; P represented the vector of new extracted features. Then, the P values wired to the classifiers that trained in supervised fashion to classify the features to the classes. The flowchart of the proposed method presented in Figure 1.

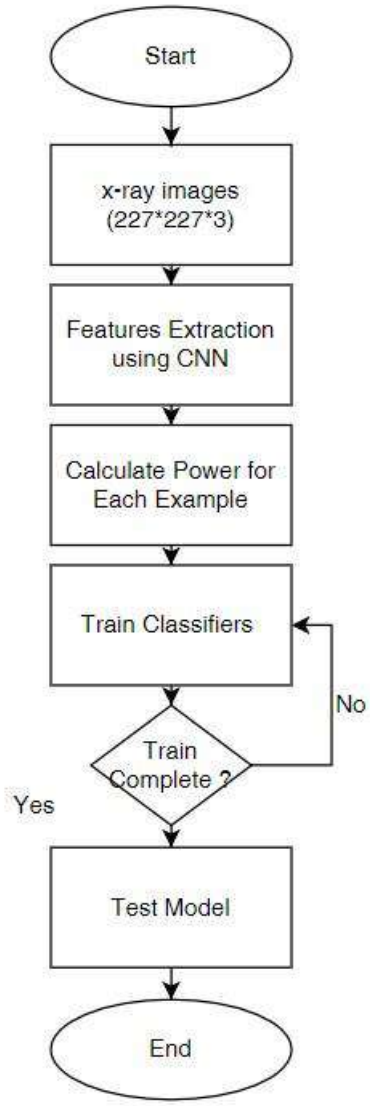


Figure -1 Flowchart of the model

Dataset

The well-known x-ray image dataset used in several state of art Covid-19 detection studies. The dataset consist from 5144 train data and 1288 test images and the whole dataset consist from 6432. The dataset consist from three classes: COVID19, PNEUMONIA, and NORMAL

3. Results

In this section, the proposed method validated in common used x-ray dataset. We next report the comparisons between different Machine Learning models and show the superior of the SVM classifier. Moreover, we show two patient level case studies along with variable importance plots demonstrating which variables the models found most important. Confusion matrix of the four classifiers: (DT, SVM, KNN, and RF) presented in the Figure 3, Figure 4, Figure 5, and Figure 6 respectively. These results are obtained after testing several iteration values to select the best iteration number. The accuracy of the DT presented in the Figure 1 according to the iterations of the training.

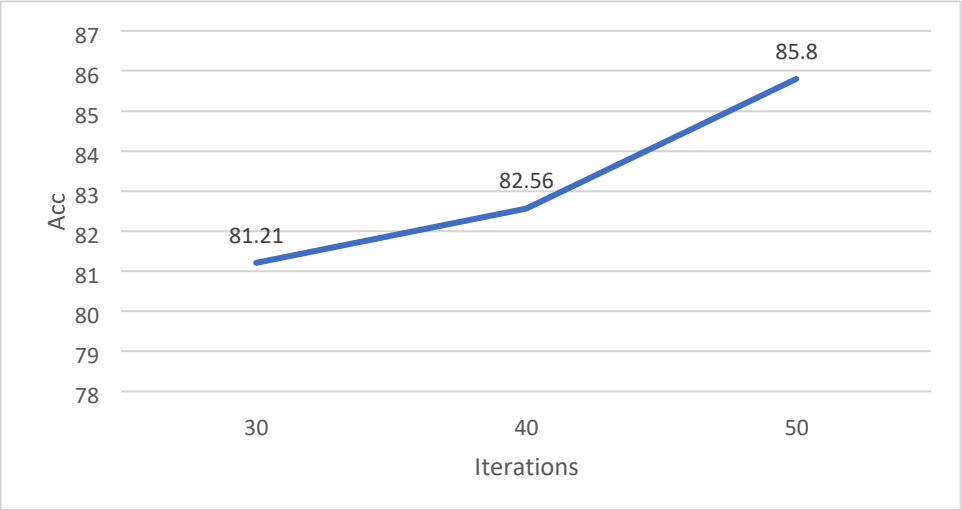


Figure -2 Iterations of CNN+ Power of Signal + DT

Furthermore, the confusion matrix of the iteration 50 which is the highest and best result presented in the Figure 3.

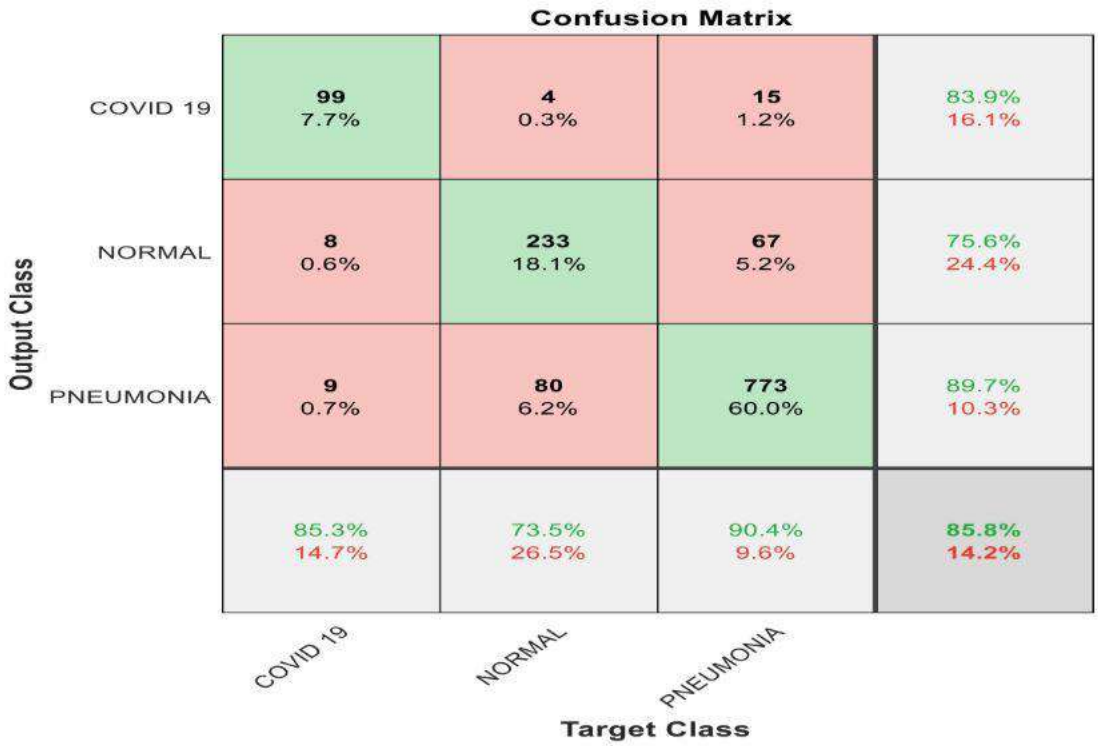


Figure -3 CNN+ Power of Signal +DT

On the other hand, the accuracies of the SVM according to the iterations presented in the Figure 4.

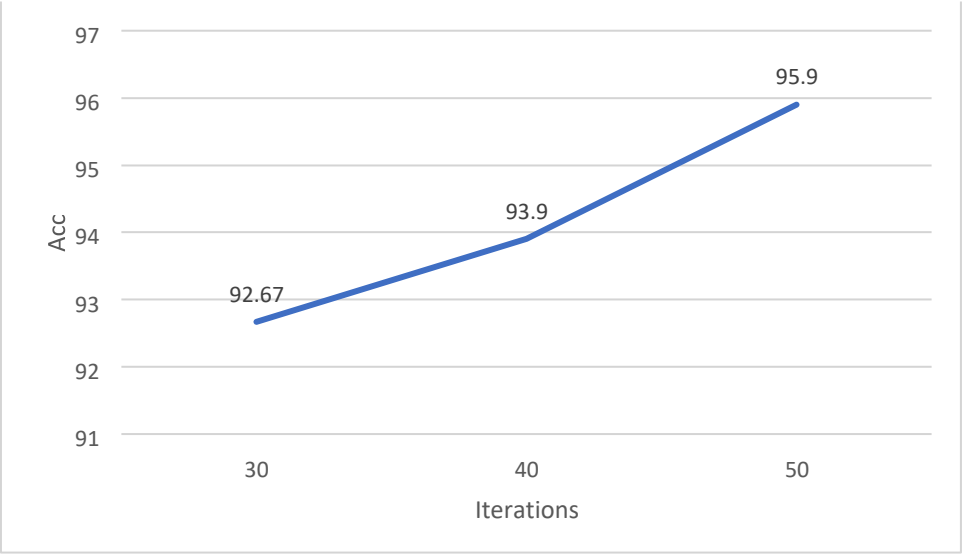


Figure -4 Iterations of CNN+ Power of Signal + SVM

Furthermore, the confusion matrix of the iteration 50 which is the highest and best result presented in the Figure 5.

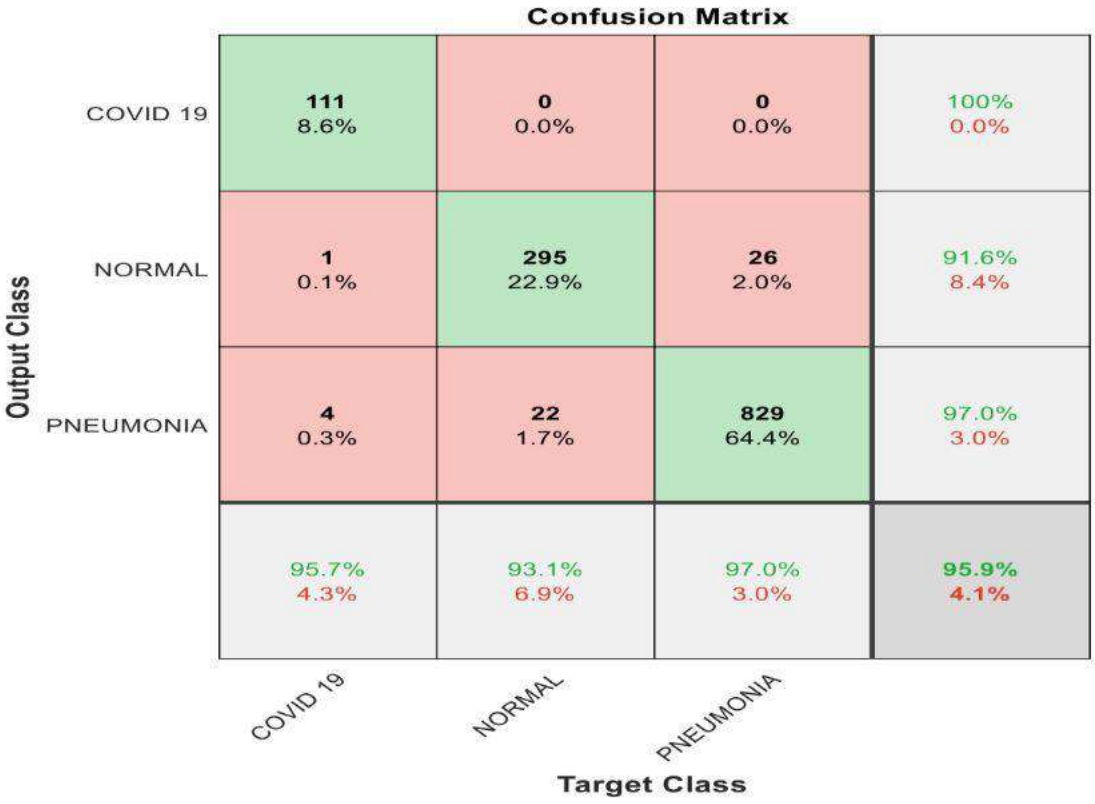


Figure -5 CNN+ Power of Signal +SVM

Moreover, the KNN iterations also presented in the Figure 6 with accuracies. The iterations that tested in these experiments are 40, 50, and 60.

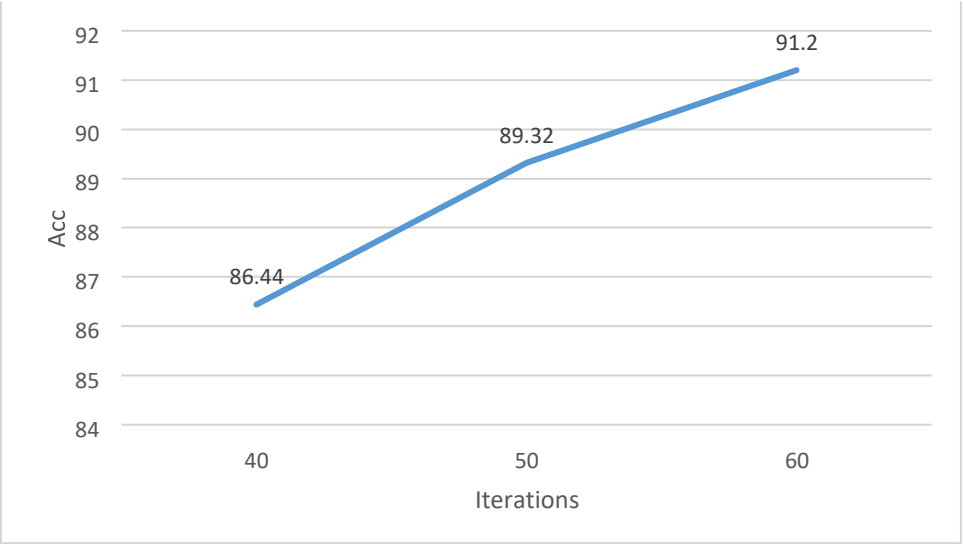


Figure -6 Iterations of CNN+ Power of Signal +KNN

Furthermore, the confusion matrix of the iteration 60 which is the highest and best result presented in the Figure 7.

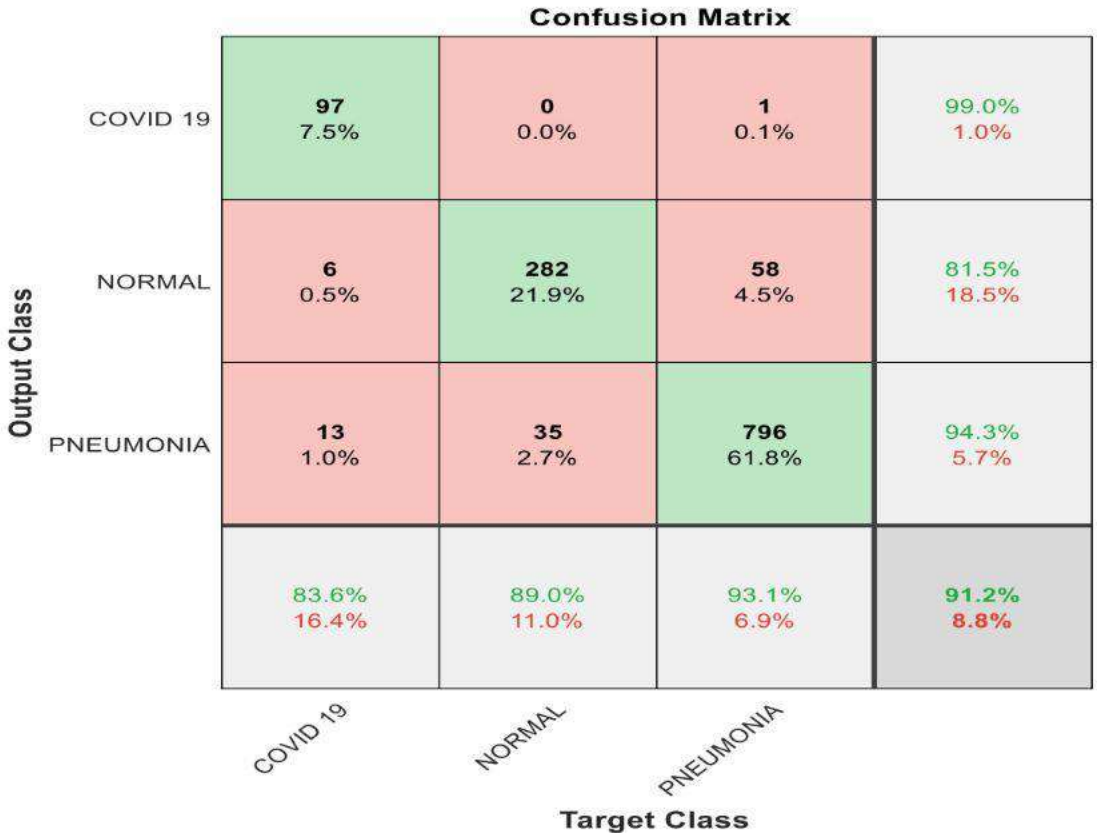


Figure -7 CNN+ Power of Signal +KNN

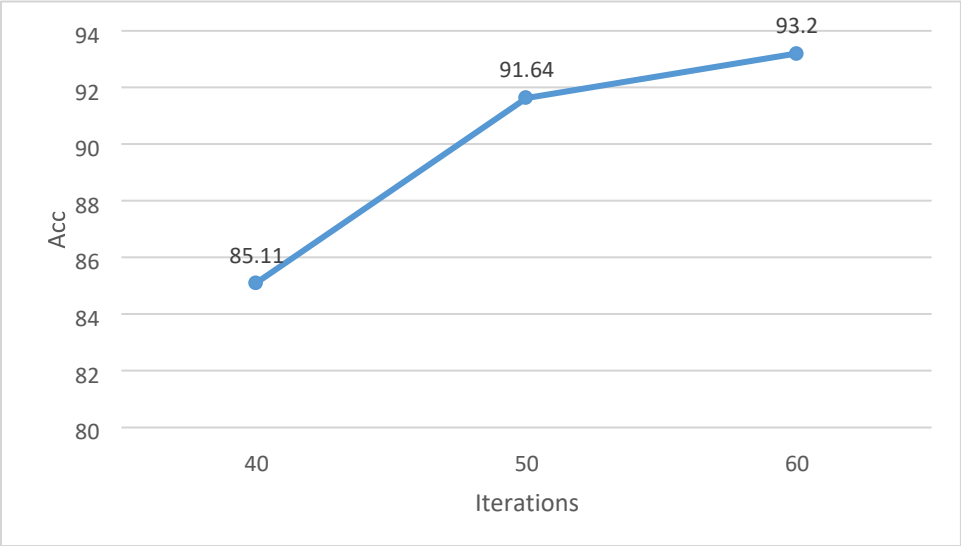


Figure -8 Iterations of CNN+ Power of Signal +RF

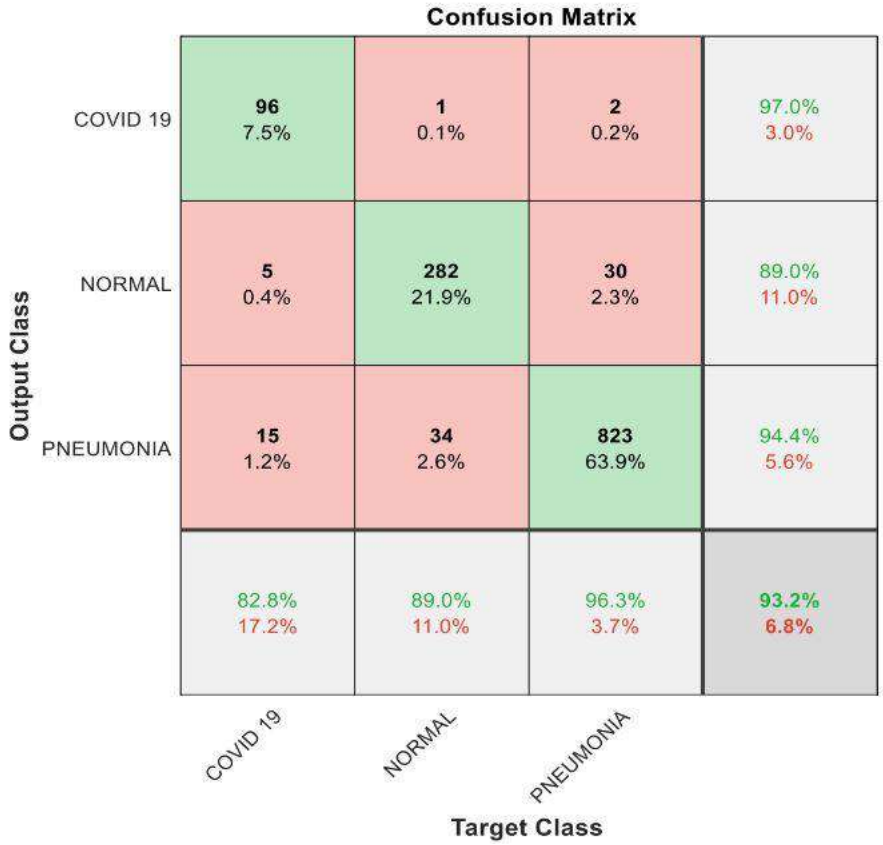


Figure -9 CNN+ Power of Signal +RF

The obtained results compared with several studies that applied deep learning techniques to the same dataset as shown in Table 1.

Table 1 Obtained Results.

Ref	Method	Acc (%)
[15]	Tailored CNN	92.3
[16]	DenseNet	88.90
[17]	Capsule Networks	95.7
[18]	DarkNet-19 based CNN	87.02
Our Method	CNN+ Power of Signal +RF	93.2
Our Method	CNN+ Power of Signal +KNN	91.2
Our Method	CNN+ Power of Signal +DT	85.8
Our Method	**CNN+ Power of Signal +SVM**	**95.9**

4. Conclusion

According to the obtained results we can conclude that the Power of signal function enhance the extracted features by the CNN. The execution of the same method without applying power of signal function leads to lose in exaction time and also presented low accuracy. The 4096 in the most time contain effective features which this leads to lose in execution time. The power of signal function is simple equation lead directly to reduce the size of features which lead to gain in the execution time.

Furthermore, in the future studies the author advises the researcher to applied other feature extracted techniques to assist CNN in the features extraction level which lead to enhance the classification accuracy and execution time. Furthermore, other transfer learning techniques can be applied in other image classification problems.

5. References

[1]. H. Tak, L. Kim, and D. K. park, "Quantum convolutional neural network for classical data classication," *Springer Quantum Mach. Intell. 4, 3,* DOI: https://doi.org/10.1007/s42484-021-00061-x (2022).

[2]. H. Li, W.Hu, W. Li, J.Li, Q. Du, and A. Plaza, "A3CLNN: Spatial, Spectral and Multiscale Attention Conv LSTM Neural Network for Multisource Remote Sensing Data Classification," IEEE © *Transactions on Neural Networks and Learning Systems,* DOI: https://doi.org/10.1109/TNNLS.2020.3028945 . (2022).

[3]. W. Fedus, J. Dean, and B. Zoph, "*arXiv A Review of Sparse Expert Models in Deep Learning, arXiv:2209.01667*, DOI: https://doi.org/10.48550/arXiv.2209.01667 (2022).

[4]. C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, P. Fan, J. Han, K. Qian, and Q. Chen, "Deep learning in optical metrology: a review" *Official journal of the CIOMP 2047-7538*, DOI: https://doi.org/10.1038/s41377-02200714-x (2022).

[5]. P. R. Matos, J. S. Neto, D. Jansen, A. G. Torre, and P. Kirchheim, "In-situ laboratory X-ray diffraction applied to assess cement hydration," *ScinceDirect Cement and Concrete Research, Volume 162, 106988, ISSN 0008-8846* DOI: https://doi.org/10.1016/j.cemconres.2022.106988 (2022).

[6]. E. Luz, P. Silva, R. Silva, L. Silva, J. Guimarães, G. Miozzo, G. Moreira, and D. Menotti, "Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images," *Springer Res. Biomed. Eng. 38, 149–162*, DOI: https://doi.org/10.1007/s42600-021-00151-6 (2021).

[7]. C. P. West, V. M. Montori, and P. Sampathkumar, "Covid-19 testing: the threat of false-negative results. In Mayo Clinic Proceeding," Mayo *Clinic Proceedings Perspectives and Controversies Volume 95, Issue 6, P1127-1129, JUNE 01*, DOI: https://doi.org/10.1016/j.mayocp.2020.04.004 ,(2020).

[8]. Y. Fang, et al., "Sensitivity of chest CT for covid-19: comparison to RT-PCR. Radiology," *Radiology Journal Vol. 296*, No. 2, DOI: https://doi.org/10.1148/radiol.2020200432 ,(2020).

[9]. Y. Yang et al., "Evaluating the accuracy of diferent respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections," *medRxivThe Preprint Server for Health Science* DOI: https://doi.org/10.1101/2020.02.11.20021493 ,(2020).

[10]. P. Wikramaratna, R. S. Paton, M. Ghafari and J. Lourenco, "Estimating false-negative detection rate of sars-cov-2 by rt-pcr," *National Library of Medicine Journal,* DOI:

https://doi.org/10.2807/15607917.ES.2020.25.50.2000568 ,(2020).

[11]. H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, and D. Shen, "Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning," IEEE *transactions on medical imaging*, DOI: https://doi.org/10.1109/TMI.2020.2992546 (2020).

[12]. L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, and K. Cao, "Artificial intelligence distinguishes COVID19 from community acquired pneumonia on chest CT," *National Library of Medicine Journal,* DOI: https://doi.org/10.1148/radiol.2020200905 .(2020).

[13]. C. Butt, J. Gill, D. Chun, B. A. Babu, "Deep learning system to screen coronavirus disease" *pneumonia*

[14]. *Applied Intelligence, Northwell Health, Article 6359 [p.]. Available from*: https://academicworks.medicine.hofstra.edu/articles/6359 .(2020).

[15]. A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks," *Computers in Biology and Medicine*, *103795* DOI: https://doi.org/10.1016/j.compbiomed.2020.103795 .(2020).

[16]. H. X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chang, K. Halsey, and J. Mei, "AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT,". *National Library of Medicine Journal*, *201491*, DOI: https://doi.org/10.1148/radiol.2020201491 (2020).

[17]. F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, and D. Shen, "Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification," *Physics in Medicine & Biology*, DOI: https://doi.org/10.1088/1361-6560/abe838 .(2020).

[18]. U. Ozkaya, S. Ozturk, and M. Barstugan, "Coronavirus (COVID-19) Classification using Deep Features Fusion and Ranking Technique," *arXiv preprint arXiv:2004.03698*, DOI: https://doi.org/10.48550/arXiv.2004.03698 (2020).

[19]. M. Z. Alom, M. M. Rahman, M. S. Nasrin, T. M. Taha, and V. K. Asari, "COVID_MTNet: COVID19 Detection with Multi-Task Deep Learning Approaches," *arXiv preprint arXiv:2004.03747*, DOI: https://doi.org/10.48550/arXiv.2004.03747 . (2020).

[20]. X. J. Yang, F. Gao, and Y. Ju, "General fractional derivatives with applications in viscoelasticity

[21]. Academic Press,"

[22]. https://www.researchgate.net/publication/330684880_General_Fractional_Derivatives_with_Applications_in_Viscoelasticity (2020).

[23]. L. Wang, A.Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images," *Sci Rep 10, 19549*, DOI: https://doi.org/10.1038/s41598-020-76550-z . (2020).