

Detection of COVID-19 infected lungs from chest x-ray using wavelet transform, HOG features extraction and SVM

Md. Hossain Ali ¹, Sadia Afrin ², Redwanul Islam ³ and Md. Rafiqul Islam ^{1,*}

¹ *Mathematics Discipline, Khulna University, Khulna 9208, Bangladesh.*

² *Department of Basic Science, Primeasia University, Dhaka, Bangladesh.*

³ *Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh.*

Global Journal of Engineering and Technology Advances, 2022, 13(02), 001–011

Publication history: Received on 05 September 2022; revised on 31 October 2022; accepted on 03 November 2022

Article DOI: <https://doi.org/10.30574/gjeta.2022.13.2.0086>

Abstract

Coronavirus SARS-CoV-2 referred to as COVID-19, is a both spreadable and infectious disease, which has footprinted a global pandemic and still infecting millions across the globe. At present, COVID-19 has made a devastating impact on our daily life. To detect coronavirus, some medical radiography technique is prominent such as chest X-ray images. This work represented the distinguishing features between normal and COVID-19 infected chest X-ray images through Discrete Wavelet transform (DWT) and Histogram of Oriented Gradients (HOG) methods which helps to indicate whether the person is COVID positive or negative. DWT and HOG transformations were performed to extract the features from the chest x-ray images. Support Vector Machine (SVM) classifier is used to the chest x-ray images for model training and validation. To evaluate the performance of the model accuracy, sensitivity, specificity and precision were calculated. DWT-SVM model provides the accuracy of 98.58%, the sensitivity of 98.38%, the specificity of 98.47% and the precision of 98.48% whereas the HOG-SVM model provides the accuracy of 99.39%, the sensitivity of 99.19%, the specificity 99.28% and precision 99.29%. So, the result indicates that the HOG-SVM model shows better performance than the DWT-SVM model. The experimental results may help the medical personnel to diagnose easily and to take the necessary steps for better treatment.

Keywords: COVID-19 Detection; Histogram Oriented Gradient; Support Vector Machine; Discrete Wavelet transform

1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) referred to as COVID-19, which is originated from the Latin word 'Coron'. Then turns into the name corona, a buzzword, which means flower crown, radiance or flower garland. Coronavirus was first detected in the 1930s and the corona-infected patient was identified in the 1960s. This 8th largest species of coronavirus, which was identified by Li Wenliang, surfaced first at Wuhan, the capital of China's Hubei region on December 30, 2019 [1]. The World Health Organization declared this coronavirus an outbreak on 7th January 2020. Indications of coronavirus contamination go in earnestness from respiratory complexity like pneumonia, kidney issue and the development of fluid in the lungs [2]. Although a chest X-ray image could serve as a practical suggestion for COVID-19, it may show some differences between normal and COVID-19 infected patients' lungs medical images. Since the COVID-19 attacks, an x-ray is utilized to identify infected and non-infected lung. X-ray images, captured by a medical practitioner, are often used to diagnose infected and non-infected chest lungs images which determines whether the patients are attacked by COVID-19 or not [3]. In the shortage of test kits, x-ray images from the gutsy population like Bangladesh are more available to identify coronavirus and necessary steps including isolation and pure appropriate treatment are quickly managed. Therefore, the experts suggest checking chest lung x-ray images to diagnose COVID-19 in the early phase [4]. Due to the lack of quick identification methods, COVID-19 disease is

* Corresponding author: Md. Rafiqul Islam
Mathematics Discipline, Khulna University, Khulna 9208, Bangladesh.

flourishing day by day [5]. Our respiratory system like the lungs is the best media where the virus can be propagated easily as a result there creates several obstacles to oxygen transmission. So quick and accurate identification of the virus is a major challenge to cut down the death rate caused by this virus for health professionals around the world [6]. Our research shows a fast, accurate and easy detection methods as such.

2. Literature Review

A significant amount of research work is carried out to identify COVID-19 viruses as follows:

Matin *et al.* [7] performed the detection through a support vector machine learning technique for COVID-19 from where the data showed an accuracy of 92.4% in classifying 154 x-ray images were confirmed that 75 are COVID-19 cases. AM. Sarhan [5] used Wavelet for 100 X-ray images and classified them by the SVM algorithm which showed an accuracy level of 94.5%. Hemdan *et al.* [1] used deep learning models (50 chest x-ray images from getting COVID-19 confirmed 25 images) to diagnose COVID-19 in x-ray images and emphasized a COVIDX-Net model which consisted of 7 CNN models and their F1 score is 89%. Amir R. [6] proposed their work for 974 x-ray images using Machine learning for classifying through SVM and deep learning-based CNN model. Their results of ML accuracy level were 80%.

Novitasari *et al.* [8] discussed a support vector machine-based system where CNN provided high accuracy level of 97.33%. The authors trained and tested the system with patient's 125 chest x-ray images were got 25 confirmed COVID-19 patients. A. Nur [2] used the HOG feature and Deep learning-based CNN model for 1584 x-ray images and they got their results which showed the accuracy of 87.34% and 93.64% respectively. Ioannis *et al.* [9] developed a deep learning model (1427 x-ray images for 224 confirmed COVID-19 images) which presented 96.78% accuracy for a two-class problem. Aras *et al.* [10] used 561 x-ray images in a deep learning-based CNN model and their research accuracy level was 99.29%. Hamza *et al.* [11] proposed an improved classification approach using wavelet and MCC features by SVM algorithm for classifying x-ray 2400 lung images and their accuracy level of 98.8%.

Lamia *et al.* [12] worked in a Wavelet-based SVM model for 714 x-ray images and their accuracy level was 95.76%. Narin *et al.* [13] got a 98% detection accuracy from 1493 chest x-ray images getting COVID-19 case 341 images through the ResNet50 model. Yash *et al.* [14] performed a Deep Learning-based model for 500 X-ray images and the accuracy level of the research was 98%. Prabira *et al.* [15] showed the ResNet50 model for X-ray images with an SVM classifier provided 95.38% accuracy. Pereira *et al.* [16] used texture features to classify chest X-ray (CX) for COVID-19 diagnosis and showed an F1 score of this model of 69%. Suresh *et al.* [17] discussed their research work through Wavelet-based and deep learning CNN model for 219 X-ray images and their Accuracy level was 99.29%.

3. Methodology

3.1. Data Collection and Description

The COVID-19 Chest x-ray dataset was collected from online hosted data [18]. As the occurrence of COVID-19 is very recent, none of the repositories contain any COVID-19 labeled data in large numbers, thereby this dataset requires relying on different sources of images. In this dataset 2460 x-ray Lung images are available where 1230 are COVID-19 infected images and 1230 are normal images. All the images are grayscale images and images are in PNG format. The visualization of x-ray images of each class is shown in figure 1.

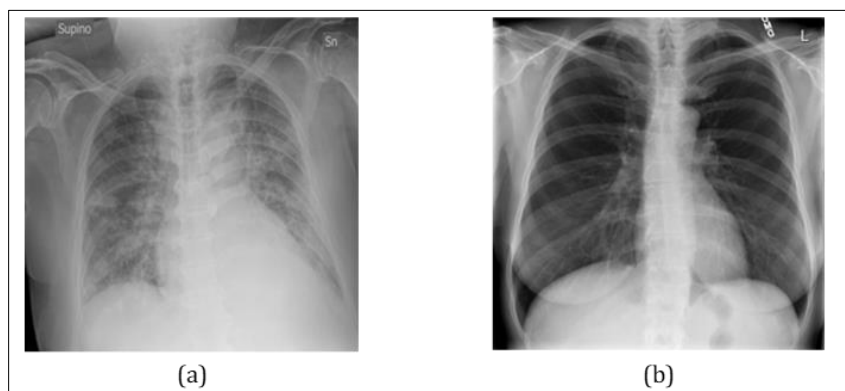


Figure 1 Sample image of (a) COVID-19 infected lung and (b) Normal lung Images

3.2. Feature Extraction Techniques

3.2.1. Discrete Wavelet Transform

In discrete wavelet transform (DWT), Symlet wavelet, Haar wavelet, Daubechies wavelet, Biorthogonal wavelet, Coiflet wavelet, and Discrete Meyer wavelet etc. are prominent for performing decomposition for extracting important features. Just to put it bluntly that the one and second-level approximation coefficients of wavelet show better results. The wavelet feature shows 4 types of co-efficient as Low-low phase filter (LL), Low-high phase filter (LH), High-low phase filter (HL) and High-high (HH) phase filter coefficients. Low-low phase filter (LL) coefficient is approximation coefficients and the High-low phase filter (HL) and High-high (HH) phase filter are details co-efficient such as horizontal details, vertical details and diagonal details co-efficient respectively [5]. The typical one-level wavelet composition of an image is shown in figure 2. DWT was applied to the chest x-ray images to perform wavelet decomposition for extracting features.

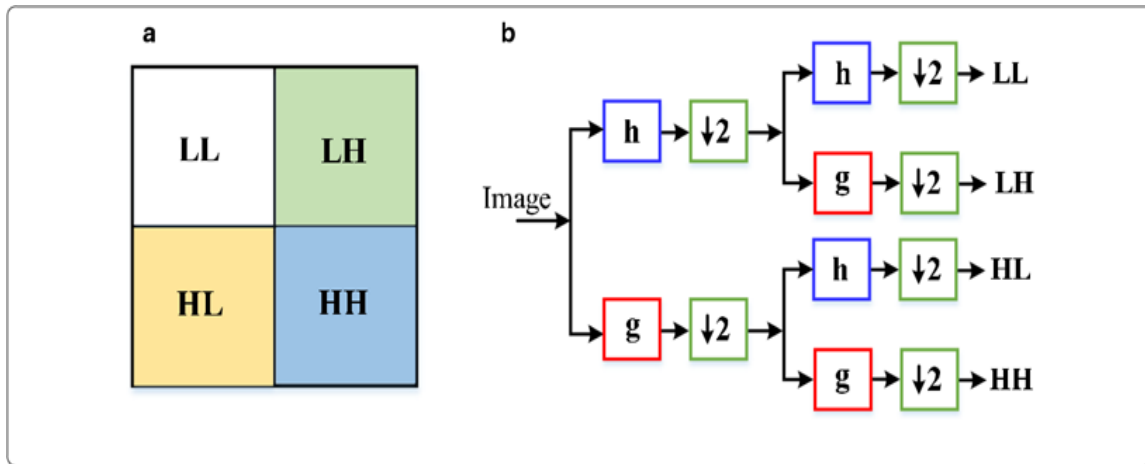


Figure 2 One Level Wavelet Decomposition

3.2.2. Implementation Steps of HOG Feature Extraction

HOG is also a well-known feature extraction technique. HOG extracts a feature that detects an object (image). The HOG feature extraction technique converts an image itself into a grayscale image and also generates this image into a gradient image. Then this feature divides into cells and overlapping blocks and also calculates the orientation histogram for each block as well as total blocks in a different region of the image using different numbers of histogram bins and becomes normalization. Then HOG shows concentrated normalization histograms [2].

3.3. Machine Learning Classifier

Support Vector Machine (SVM) is a well-known binary classifier. It is considered a powerful tool for real data classification. SVM is a discriminative classifier that locates a separating hyperplane between two classes which will identify the largest minimum margin between them shown in Figure 3. The advantages of support vector machines are: effective in cases where the number of dimensions is greater than the number of samples, using a subset of training points in the decision function makes it more memory efficient and different kernel functions can be specified for the decision function [6].

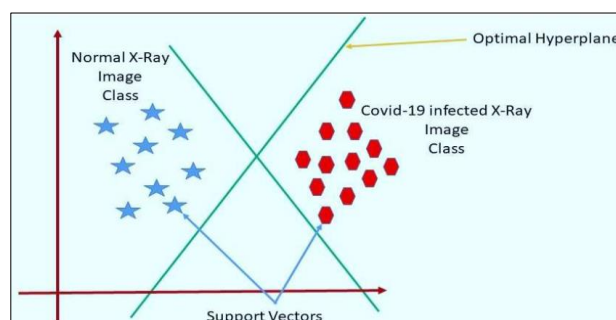


Figure 3 Support Vector Machine [6]

3.4. Proposed Approach

Raw x-ray images were first given to the model as input. In the preprocessing part, all images were resized to 224*224 pixels and smoothing, rotating, flipping, and normalizing the x-ray image for data augmentation purposes. After that, features were extracted from x-ray images using DWT and HOG transformation techniques. A new features dataset was constructed using the extracted features from the chest x-ray images. Then the dataset was divided into two parts: the training dataset and the testing dataset. We use 80% data of our dataset to train our model, and the remaining 20% of the data is considered to validate the trained model. To train our model SVM classifier was used. Then, hyperparameter tuning was performed to optimize the performance of the classifier model. Kernels, regularization and gamma are the main three parameters of the SVM classifier. The main function of the kernel is to take low-dimensional input space and transform it into a higher-dimensional space. Regularization (C) is the penalty parameter, which represents misclassification or error term. The misclassification or error term tells the SVM optimization how much error is bearable. This is how SVM can control the trade-off between decision boundary and misclassification terms. Gamma defines how far influences the calculation of a plausible line of separation. Here, radial basis function (RBF) was used as kernel, 10 was used as regularization value and gamma value 1 was used in this work. Finally, the performance of the model was evaluated using an evaluation matrix. All analyses were performed using MATLAB R2018b in an Intel Core i7 6500U processor without any graphics processing unit. Figure 4 represents the schematic diagram of the proposed workflow.

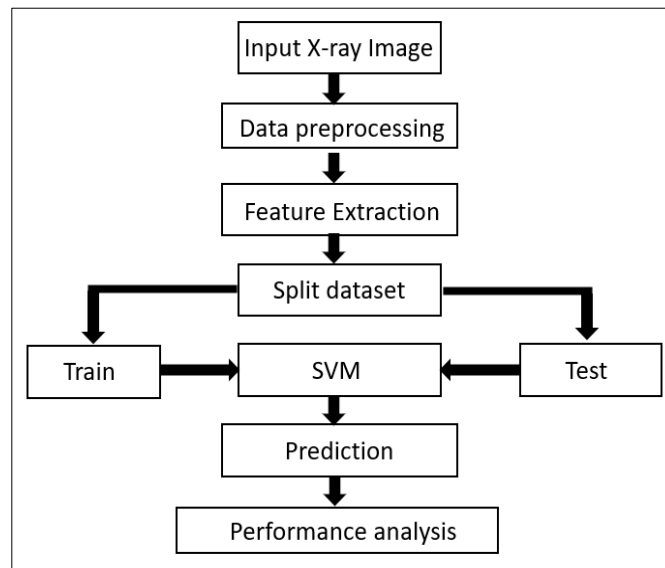


Figure 4 Schematic diagram of the methodology

3.5. Evaluation Matrix

The test performance of the work is associated with such classification images in terms of accuracy analysis, sensitivity analysis, specificity analysis and precision analysis which is calculated using equation (i)-(iv).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (i)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \dots\dots\dots (ii)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \dots\dots\dots (iii)$$

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots (iv)$$

In the above equations, TP, TN, FP, and FN are named for true positive, true negative, false positive and false negative rates. Accuracy means correctly classifying cases divided by the total cases [19]. Precision means the fraction of accurately identified instances in actually positive cases. Sensitivity calculates the ratio of accurately differentiated positive samples on the other hand specificity means the ratio of accurately distinguished negative samples [20].

Though the accuracy rate is the prime target of several classification algorithms, the accuracy rate may be attained by being involved in certain cases.

4. Results and Discussion

4.1. Image Decomposition Through DWT

Chest x-ray images were given as input to wavelet transform to get rows and columns co-efficient. Figure 5 shows LL, LH, HL and HH remarks as an approximation, horizontal details, vertical details and diagonal details co-efficient respectively.

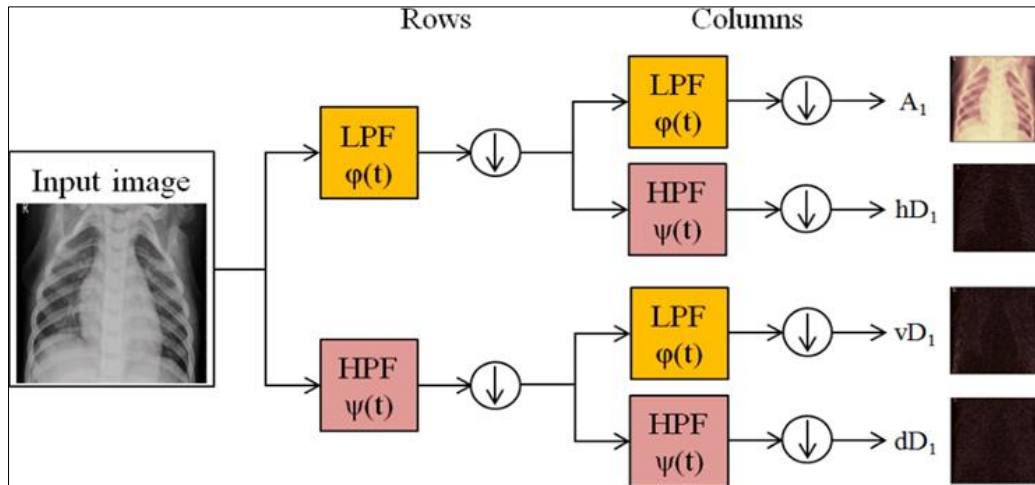


Figure 5 One-level decomposition of the chest x-ray image through DWT

4.1.1. Histogram Density Analysis in Different Wavelet Decomposition Levels

A chest x-ray image was given as the input of the model. First, using DWT the input image was decomposed into level-1 and level-2. From these two types of decomposition, some coefficient was measured. This coefficient was considered as the feature of the image. Using this coefficient histogram was constructed.

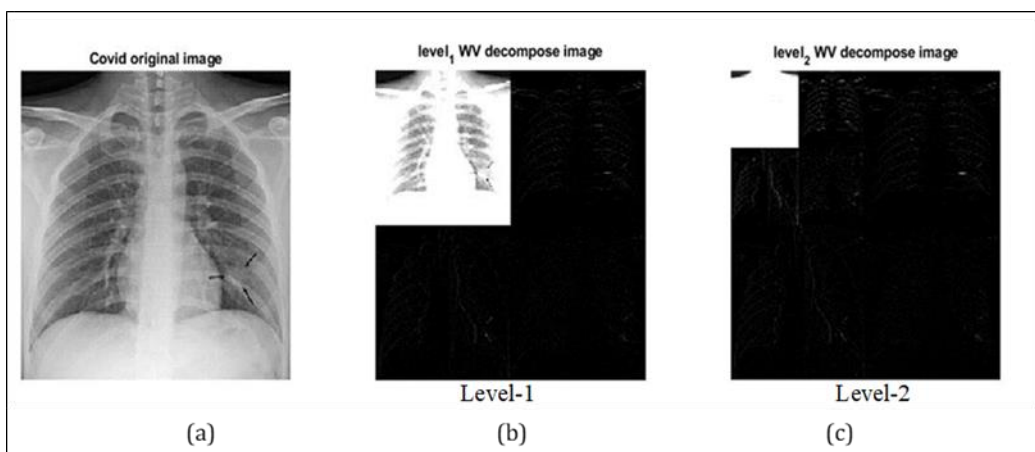


Figure 6 (a) COVID infected image, (b, c) level-1 and level-2 decomposition through DWT for COVID-19 image

A COVID-19 infected lung image decomposition is shown in figure 6, where 6(a) is the input x-ray image, 6(b) and 6(c) shows the level-1 and level-2 decomposition. Discrete Wavelet Decomposition shows level-1 and level-2 as approximation co-efficient and details co-efficient which is shown in figure 7. Comparing the coefficients, it is seen that only the approximation part always carries the main features of the image. We also observed that the histogram density for level-1 is up to 3000 cm and for level-2 is up to 1000 cm.

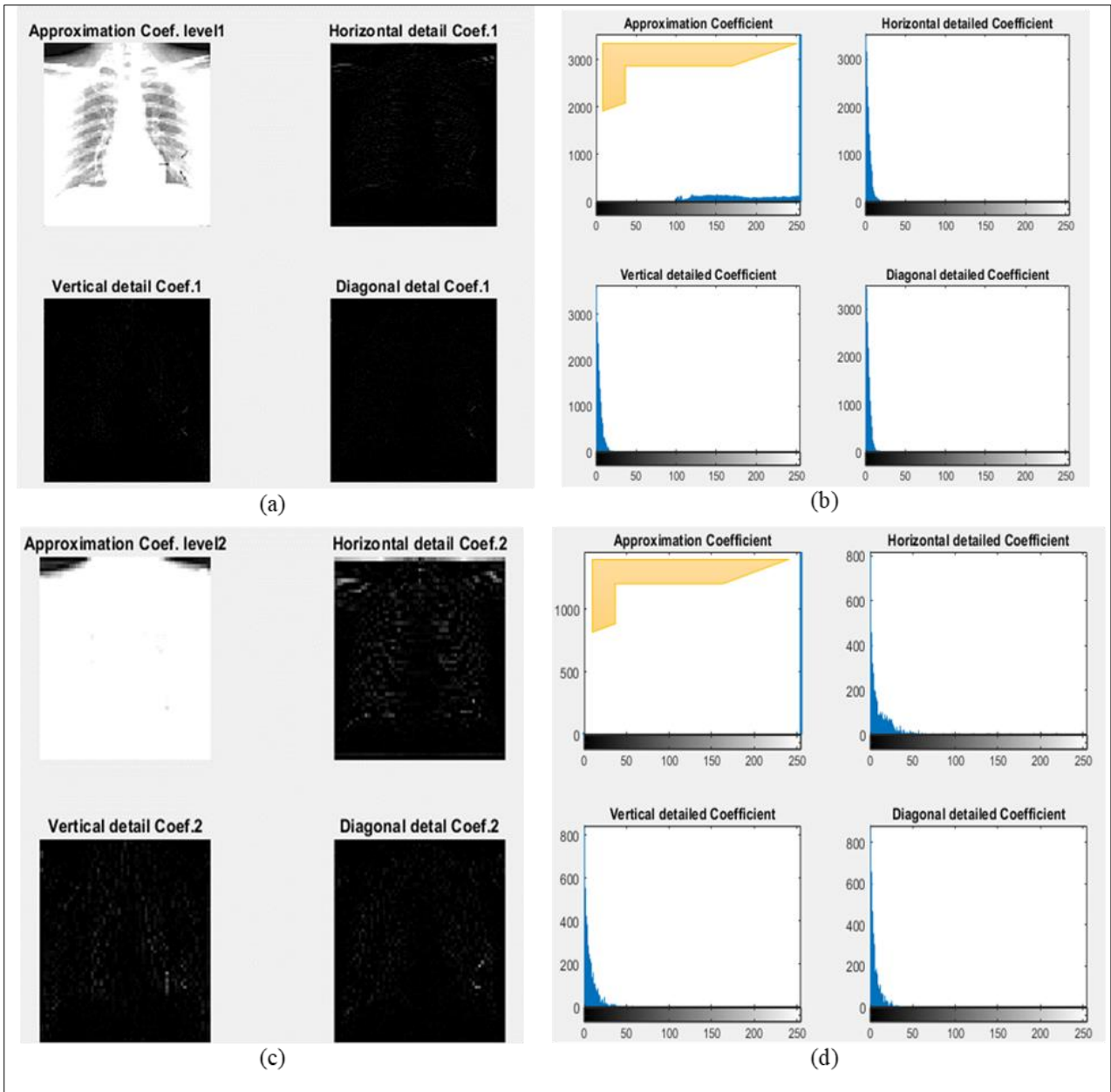


Figure 7 For COVID-19 infected lung image (a, b) level-1 decomposition and its corresponding coefficient represented in the histogram, (c, d) level-2 decomposition and its corresponding coefficient represented in the histogram

A normal lung image decomposition is shown in figure 8, where 8(a) is the input x-ray image of a normal lung, 8(b) and 8(c) shows the level-1 and level-2 decomposition of that image. Level-1 and level-2 are again divided by approximation co-efficient and details co-efficient. The approximation coefficient part always carries the main features. We also observe that the histogram density for level-1 is up to 2500 cm and for level-2 is up to 1000 cm for a normal lung.

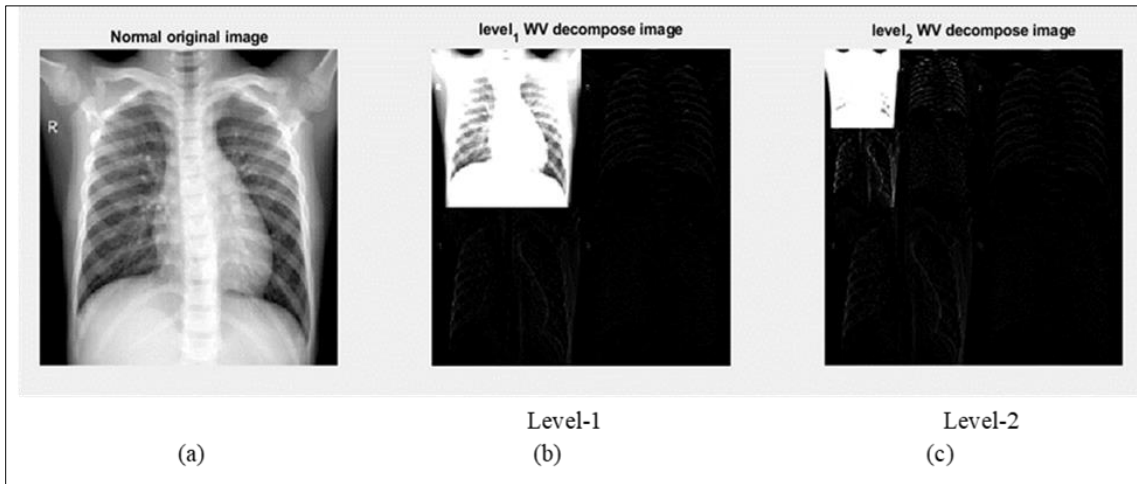


Figure 8 (a) Normal lung image, (b, c) level-1 and level-2 decomposition through DWT for Normal lung image

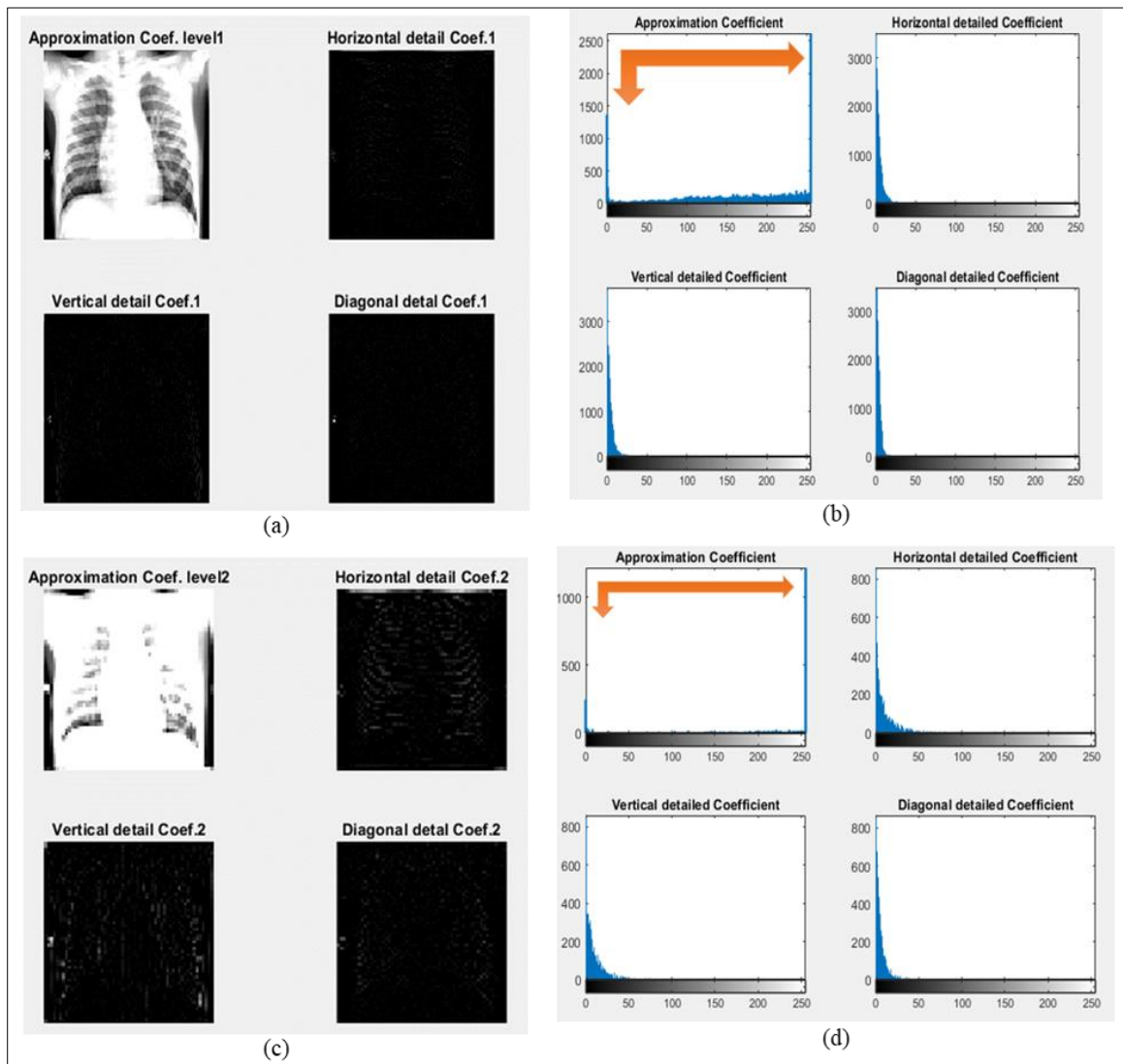


Figure 9 For normal lung image (a, b) level-1 decomposition and its corresponding coefficient represented in the histogram, (c, d) level-2 decomposition and its corresponding coefficient represented in the histogram

4.2. Visualization and Feature Extraction by HOG Method

In the HOG method, HOG features were extracted from the input x-ray image. Figure 10 and figure 11 represented the input image, HOG features and visualization of that HOG features. HOG features of the infected and normal lung is differentiable. In the visualization stage, a shade of the COVID-19 infected image presented in Figure 7(c) is nominal and the shade of the normal x-ray image presented in Figure 8(c) is highly dense. These highly distinguishable HOG features were given as the input of the SVM model.

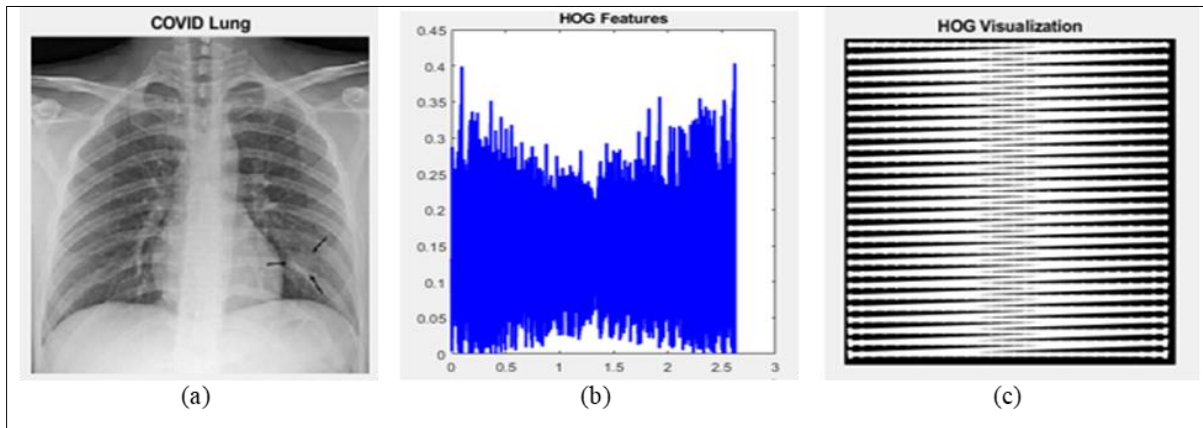


Figure 10 Represents (a) Infected COVID-19 X-ray image, (b) HOG features and (c) HOG feature visualization

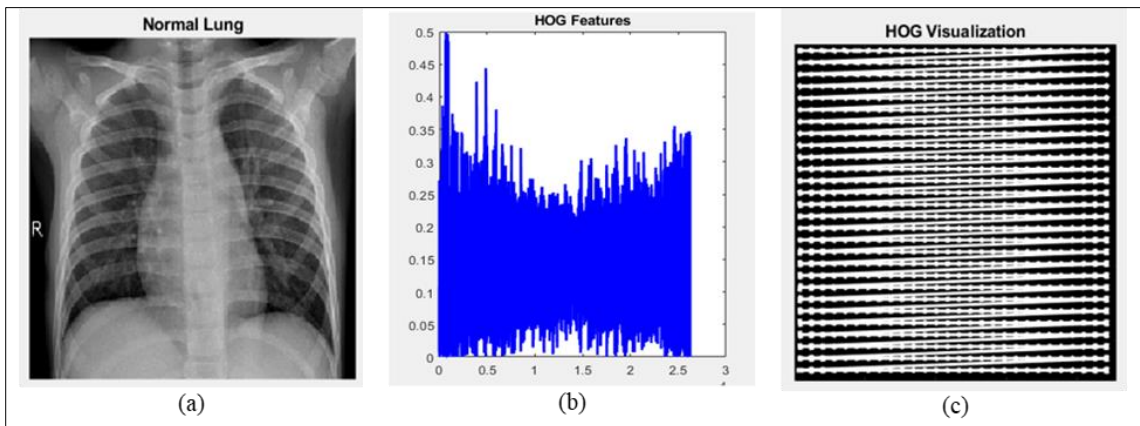


Figure 11 Represents (a) Normal lung X-ray image, (b) HOG features and (c) HOG feature visualization

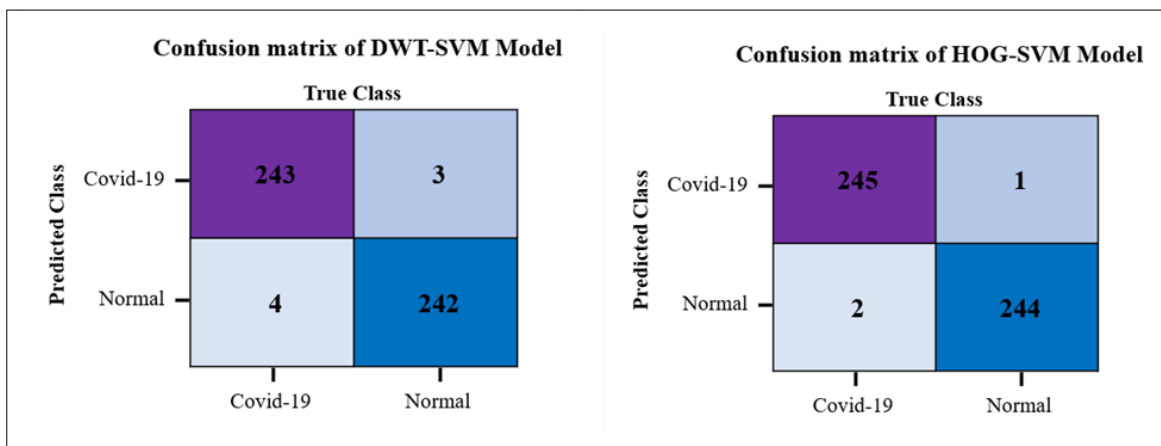


Figure 12 Confusion Matrix of DWT-SVM model and HOG-SVM model

For the performance analysis confusion matrix is used. The confusion matrix was calculated using the testing dataset. The confusion matrix generally compares the predicted result with the true result. The confusion matrix of DWT-SVM and HOG-SVM has shown in figure 12.

To minimize the probability of pandemic consequences, researchers especially focus on clinical studies to get good quality of information about COVID-19 using chest x-ray images. In the vein of this study, the DWT-SVM model and HOG-SVM model were performed to detect and automatically classify chest x-ray images. DWT-SVM model provides the accuracy of 98.58%, the sensitivity of 98.38%, the specificity of 98.47% and the precision of 98.48% whereas the HOG-SVM model provides the accuracy of 99.39% the sensitivity of 99.19%, the specificity 99.28% and precision 99.29%. So, suffice it to say that HOG-SVM is more accurate than the DWT-SVM model. The experimental results of the work are presented in the following chart in Table 1.

Table 1 Performance of the two models

Model	Accuracy	Sensitivity	Specificity	Precision
DWT-SVM	98.58%	98.38%	98.47%	98.48%
HOG-SVM	99.39%	99.19%	99.28%	99.29%

A comparison between the existing model and our model in terms of accuracy has been shown in Table 2. From the comparison, we may say that our model shows more accuracy and it is also quicker than other approaches.

Table 2 Compare of our model with different models in terms of accuracy

Reference	Model	Accuracy
A. Nur [2]	HOG + CNN	98.36%
AM. Sarhan <i>et al</i> [4]	Wavelet + SVM	94.5%
M. Aras [9]	Wavelet + CNN	96.07%
M.E.H. Chowdhury <i>et al</i> [18]	CheXNet	97.74%
F. Ucar <i>et al</i> [21]	Bayes-SqueezeNet	98.30%
M. F. Aslan <i>et al</i> [22]	CNN-BiLSTM	98.70%
T. Ozturk <i>et al</i> [23]	DarkCOVIDNet	87.02%
This work	DWT + SVM	98.58%
	HOG + SVM	99.39%

5. Conclusion

The healthcare systems of many countries have been devastated by the dreadful coronavirus pandemic that's why it is necessary to detect COVID-19 in a quicker, easier, and cheaper way that can help in saving lives and shorten the burden on healthcare costs. So, machine learning models can play a vital role in identifying COVID-19 infected patients in their early stages. Our research work is mainly based on machine learning conducted by Discrete Wavelet Transform and Histogram of Oriented Gradient feature. The accuracy of the DWT-SVM model and HOG-SVM model are 98.58% and 99.39% respectively. For the same x-ray images, the HOG-SVM model shows more accurate results to identify COVID-19 infected lungs than the DWT-SVM model. Besides, after comparing our model with other architectural models, we have observed that the HOG-SVM model shows more accuracy than other existing models. So, it is heartening to say that our model is more convenient to any other existing model.

Compliance with Ethical Standards

Acknowledgments

The authors wish to acknowledge the Mathematics Discipline, Khulna University, Khulna, Bangladesh for facilitating the work.

Disclosure of conflict of interest

The Authors confirm that the content of this manuscript has no conflict of interest.

References

- [1] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images, arXiv [eess.IV], 2020. [Online]. Available: <http://arxiv.org/abs/2003.11055>
- [2] A. Nur, COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning, *Sensors MDPI*, vol. 21, pp. 1480–1488, 2021.
- [3] T. I. Palash, R. Islam, M. Basak, and A. Dutta Roy, Automatic classification of COVID-19 from chest X-ray image using convolutional neural network, in *2021 5th International Conference on Electrical Information and Communication Technology (EICT)*, 2021.
- [4] T. I. Palash, R. Islam, and S. Debnath, Low-cost smart COVID-19 patient monitoring and support system, in *2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 2021.
- [5] AM. Sarhan, Detection of COVID-19 Cases in chest X-ray images using wavelet and support vector machines, *Science and Business Media*, vol. 8, no. 1, pp. 5–13, 2021.
- [6] R. Amir, COVID-19 Detection Empowered with Machine Learning and Deep Learning Techniques: A Systematic Review, *Applied Science*, vol. 9, no. 2, pp. 11–21, 2021.
- [7] A. Matin, A. Rahman, S M A Shuaeb, A. Hossen. Automatic COVID-19 infected chest X-ray images classification using support vector machine, *International Journal of Trend in Scientific Research and Development*, 2021; 5:4.
- [8] D. C. R. Novitasari et al., Detection of COVID-19 chest X-ray using support vector machine and convolutional neural network, *Commun. Math. Biol. Neurosci*, p. Article-ID, 2020.
- [9] I. D. Apostolopoulos and T. A. Mpesiana, COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, *Phys Eng Sci Med*, vol. 43, no. 2, pp. 635–640, 2020.
- [10] M. Aras, The investigation of multiresolution approaches for chest X-ray image based COVID 19 detection, *Health Information and Science System*, vol. 8, pp. 29–38, 2020.
- [11] H. A. Owida, A. Al-Ghraibah, and M. Altayeb, Classification of Chest X-Ray Images using Wavelet and MFCC Features and Support Vector Machine Classifier. *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7296–7301, 2021.
- [12] L. N. Mahdy, K. A. Ezzat, H. H. Elmousalami, H. A. Ella, and A. E. Hassanien, Automatic X-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine, *bioRxiv*, 2020.
- [13] A. Narin, C. Kaya, and Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks, *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, 2021.
- [14] Y. Karbhari, A. Basu, Z.-W. Geem, G.-T. Han, and R. Sarkar, Generation of synthetic chest X-ray images and detection of COVID-19: A deep learning based approach, *Diagnostics (Basel)*, vol. 11, no. 5, p. 895, 2021.
- [15] P. K. Sathy and S. K. Behera, Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis, *Multimed. Tools Appl.*, vol. 81, no. 7, pp. 9631–9643, 2022.
- [16] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. G. Costa, COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios, *Comput. Methods Programs Biomed.*, vol. 194, no. 105532, p. 105532, 2020.
- [17] G. Suresh, Wavelet Based CNN for Diagnosis of COVID 19 using Chest X Ray, *Materials Science and Engineering*, pp. 1084–1088, 2021.

- [18] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, Can AI help in screening viral and COVID-19 pneumonia?, IEEE Access, vol. 8, pp. 132665-132676, 2020.
- [19] R. Islam, S. Debnath, and T. I. Palash, Predictive analysis for risk of stroke using machine learning techniques, in 2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), 2021.
- [20] R. Raen, M. M. Islam, and R. Islam, Diagnosis of retinal diseases by classifying lesions in retinal layers using a modified ResNet architecture, in 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), 2022.
- [21] F. Ucar and D. Korkmaz, COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images, Med. Hypotheses, vol. 140, no. 109761, p. 109761, 2020.
- [22] M. F. Aslan, M. F. Unlarsen, K. Sabanci, and A. Durdu, CNN-based transfer learning-BiLSTM network: A novel approach for COVID-19 infection detection, Appl. Soft Comput., vol. 98, no. 106912, p. 106912, 2021.
- [23] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med., vol. 121, no. 103792, p. 103792, 2020.