Cihan University-Erbil Scientific Journal (CUESJ)

Research Article

Automatic Classification of Coronavirus Disease-19 Chest X-Ray Images Using Local Binary Pattern and Binary Particle Swarm Optimization for Feature Selection

Bazhdar N. Mohammed[1], Firas H. Al-Mukhtar[2], Raghad Z. Yousif[1], Yazen S. Almashhadani[3]

[1]Department of Physics, Salahaddin University-Erbil, Iraq, [2]Department of Information Technology and Computer Science, Catholic University in Erbil, Iraq, [3]Department of Communication and Computer Engineering, Cihan University-Erbil, Kurdistan Region, Iraq

ABSTRACT

Novel Coronavirus disease 2019 (COVID-19) is a type of pandemic viruses that causes respiratory tract infection in humans. The clinical imaging of Chest X-Ray (CXR) by Computer-Aided Diagnosis (CAD) plays an essential role in identifying the patients infected by COVID-19. The objective of this paper presents a CAD method for automatically classifying 110 frontal CXR images of contagious people according to Normal and COVID-19 infection. The proposed method contains four phases: Image enhancement, feature extraction, feature selection, and classification. Gaussian filter is performed to de-noise the images and Adaptive Histogram Equalization for image enhancement in the pre-processing step for better decision-making. Local Binary Pattern features set are extracted from the dataset. Binary Particle Swarm Optimization is considered to select the clinically relevant features and develop a robust model. The successive features are fed to Support Vector Machine and K-Nearest Neighbor classifiers. The experimental results show that the system robustness in classification COVID-19 from normal images with average accuracy 94.6%, sensitivity 96.2%, and specificity 93%.

Keywords: Computer-Aided Diagnosis, chest-x-ray, coronavirus disease 2019, local binary pattern, binary particle swarm optimization, classification

INTRODUCTION

At the end of December 2019, a new type of worldwide outbreak, Coronavirus: The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is called novel coronavirus (Coronavirus disease 2019 [COVID-19]) has been announced in Wuhan, China. After that, it was declared a global health emergency by the Centre of Disease Control and Prevention (CDC) and the World Health Organization (WHO).[1] Since the past two decades, other known case of a typical pneumonia SARS-COV has been discovered originating from a healthcare worker in Guangdong Province, China.[2] SARS-COV was a relatively rare epidemic disease; the incidence killed approximately 800 persons among 8437 infected cases which belongs to 11% case of fatality rate.[3] Furthermore, 9 years after Guangdong's plague, another type of this virus is Middle East Respiratory Syndrome (MERS) identified in Saudi Arabia in 2012, also known as camel flu, because it is believed this kind originated from the camel then transported to humans by contacting.[4] During the tests, many differences and similarities in the symptoms, screening features, and management of Corona disease types have been identified. The rapid rate in spreading of COVID-19 is considered one of the main diversities of nowadays novel Coronavirus.[5] Nowadays, many radiological images are captured in medical centers and hospitals. Still, defects in diagnostic tools are one of the major problems that all countries face during the initial stages of this pandemic disease. For these reasons, fast, reliable screening of COVID-19 is very crucial and impressive. Besides therapeutic and clinical guides, imaging provides extra help to the physicians, and it helps to control the disease from cross-infection effectively. The American College of Radiology suggests that Chest X-Ray (CXR) may be advised to reduce the risk of spreading.[6] Computed Tomography (CT)/CXR tool for imaging is one of the key components besides the

Corresponding Author:
Bazhdar N. Mohammed, Department of Physics, Salahaddin University-Erbil, Iraq.
E-mail: bazhdar.sh.mohammed@su.edu.krd

Received: May 16, 2021
Accepted: October 18, 2021
Published: November 10, 2021

DOI: 10.24086/cuesj.v5n2y2021.pp46-51

clinical assessments that may help early screening suspected pneumonia cases.

Moreover, CXR is a convenient and low-cost modality for clinical screening and follow-up many pulmonary diseases.[7] The images of various viral pneumonia are similar and they overlap with other infectious during exposure dose that makes a poor contrast. As a result, it may not be easy for radiologists to differentiate COVID-19 from normal chest images in some screening cases that may lead to make a wrong decision about the patient's situation.[8] Throughout this pandemic, many groups and researchers have reported different systems of COVID-19 pneumonia detection by using Computed Aided Diagnosis (CAD). The CAD consists of pre-processing, features extraction and selection, classification, and post-processing techniques. The main purpose of this work-study is to develop a CAD program for screening and classifying CT/CXR images according to Normal and COVID-19 on a small data set that will be useful in decreasing the workload of healthcare personnel. The contributions provided to this study is (1) implementation diagnosis system for enhancement the images using Gaussian filter, (2) the usage of Local Binary Pattern (LBP) to extract the texture features from the images, and (3) performing the classification technique using K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) classifier for obtaining suitable Accuracy, Sensitivity, Specificity, and Precision. During the spreading of COVID-19 as a pandemic disease till now, researchers have tried to use different machine learning techniques to classify the patient's diagnostic tests based on radiological images that assist the doctors for accurate clinical decision which reduced the number of deaths. Abbas *et al*. suggested a method for classifying COVID-19 images in XRD diagnostic images by performing a transfer algorithm. The feature vector of datasets is classified based on Convolution Neural Network (CNN) technique.[9] Another intelligence system was proposed by Purohit *et al*.[10] to separate COVID-19 infected X-ray and CT scan images of chest based on k-mean clustering as an unsupervised method for feature extraction; then the images are classified by CNN machine learning technique. To identify COVID-19 cases from normal images Narin *et al*.[11] proposed a method Deep CNN. For this purpose, the dataset was divided into two parts: 80% for the training phase and 20% for the testing phase based on 5-fold cross validation. In Elasnaoui *et al*.[12] The authors presented a deep learning model that allowed to detect COVID-19 from healthy people using CXR images using recent deep learning methods.

METHOD AND MATERIALS

The proposed methodology for the COVID_19 CXR image diagnosis system passes through the fundamental steps in machine learning data acquisition, image pre-processing, feature extraction, feature selection, and classification steps. The proposed method is shown in Figure 1.

Data Acquisition

Data acquisition was retrieved from an open dataset of COVID-19 cases with CT/CXR. The database is already publicly available in GitHub and it includes complete mailing addresses, telephone numbers, and e-mail addresses. This document represents the stage-two template.[8] This collaborative open data source allows radiologists and researchers to develop a machine learning model for COVID-19. For this work, 110 digital radiographic images were used and equally distributed on normal and COVID-19 cases.

Pre-processing

In the pre-processing step, the data were prepared to maximize the diversity between the intensity of pixels and improve the accuracy of classification results.[13] Here, at first, the images were resized to (576 × 576) to reduce the processing time and a Gaussian filter was performed to minimize the noise; then, Adaptive Histogram Equalization was applied to increase the contrast and image qualification improvement.

Feature Extraction

Feature extraction plays an essential role in any classification task. Medical images of different categories usually appeared by three matrices: Object, edge, and background, which can be distinguished with their feature characteristics.[14] If the features are precisely selected, the classification result will provide helpful information and pattern to distinguish between the normal and COVID-19 images. LBP is an effective set of texture descriptor that thresholds the intensity of neighboring pixels based on the value of the central pixel.[15] Having a value equal or greater than the threshold pixel will assign the bit 1 and the remaining others whose values are smaller than the central pixel value will have bit 0. A bit vector is formed by restricting all these binary digits in a clockwise motion. The decimal point of the central pixel can be achieved by converting the binary sequences. Mathematically, the LBP for every pixel is represented as: [16]

$$LBR(P,R)=\sum_{p=0}^{p-1} f(g_p - g_c)2^p \tag{1}$$

Where P denotes the total number of pixels surrounding the central pixel in radius R. Furthermore, g_p, and g_c represent the intensity of central pixel and neighbor pixels, respectively. In the LBP feature vector, each central pixel in (x,y) image is surrounded by eight neighbors if the radius of circulation (R) equals one. At the same time, the central value is equal to the threshold value of any neighbor pixel.

Feature Selection

The role of feature selection is to simplify the model by making a vector with relevant features, shorter training time, and reducing the over fitting case because a real dataset contains statistically irrelevant features that may create a problem in classification tasks.[17,18] Nowadays, the Binary Particle Swarm Optimization (BPSO) metaheuristic algorithm is considered as an effective modern technique for feature selection with less running consumption Figure 2.[19]

It works based on a set of candidate solutions of randomly generated particles. The particles are moving around the search space to find the global best-optimized solution due to mathematical formula over both particle's position x and velocity v. According to the used objective function, global best (g_{best}) and local best (l_{best}) of moving particles will be initiated in regards to the optimization problem. Then, after

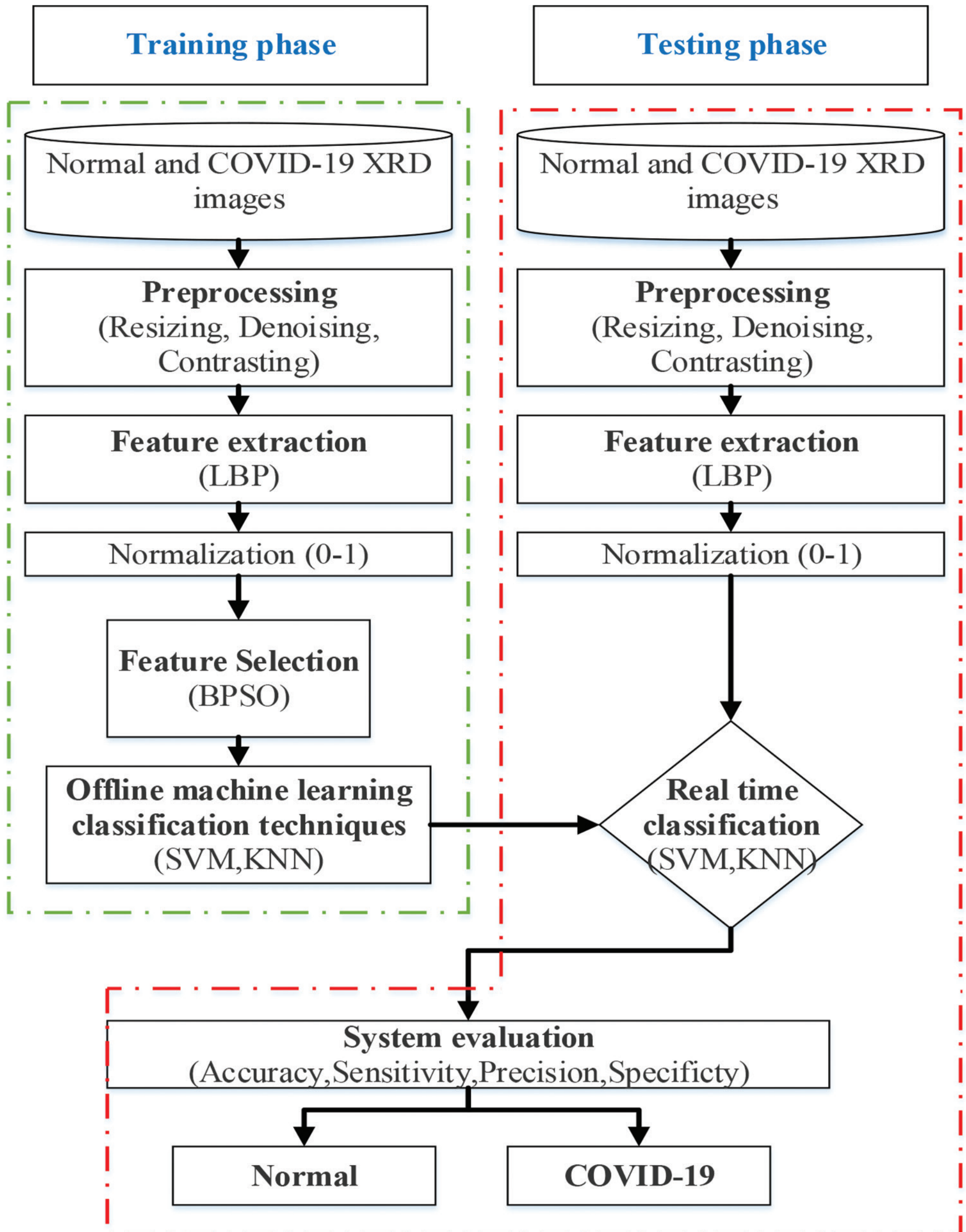

Figure 1: Flow chart of proposed methodology

each iteration, the particles try to update their positions and velocities according to the particle's last destination and best moving, which the best candidate records until the searching process is completed.[20] The next position and velocity for each particle can be determined by:

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (2)$$

$$v_{id}^{t+1} = w * v_{id}^{t} + c1 * r1(p_{id} - x_{id}^{t}) + c2 * r2(p_{gd} - x_{id}^{t}) \quad (3)$$

Where P_{id} and P_{gd} represent l_{best} and g_{best}, respectively, in the d^{th} dimension. c1 and c2 are acceleration constants. r1 and r2 are random change values between 0 and 1. w is inertia weight to set up a balance between previous and current velocity, which refers to exploitation and exploration [Table 1]. With defining the construction factor (K), the update of eq. (3) becomes:

$$v_{id}^{t+1} = K * v_{id}^{t} + c1 * r1(p_{id} - x_{id}^{t}) + c2 * r2(p_{gd} - x_{id}^{t}) \quad (4)$$

Where $K = \frac{2}{2 - \phi - \sqrt{\phi^2 - 4}}$ and $\phi = c1 + c2$, $c1 + c2 < 4$ (5)

Under the equation's balance, mathematically equation w from eq. (3) is equal. Once the relevant features were selected, passed features were labeled and split into training and testing sets based on a 10-fold cross-validation strategy. Class 0 refers to the normal images and Class 1 determines the presence of COVID-19 disease.

CLASSIFICATION

Image classification is an important step for observing whether the model can classify infected images from normal images or not. Two classification methods SVM and KNN were performed on training and testing datasets. SVM classifier determines the best hyperplane to distinguish positive and negative training samples.[21] It can project the low dimensional input data in a higher dimensional feature by choosing the proper kernel function that separates cases of class labels linearly in multidimensional space.[22] KNN captures the idea of similarity with some mathematics, we might have learned in calculating the distance between the points on the graph. The selection of a suitable value of distance (K) makes the model's performance better.[23] The confusion matrix was used to better describe the gained classification models and understand which data classes are most often misplaced when it makes a prediction. Each column of the matrix corresponds to a predicted classes and the row represents the actual classes. The number of incorrect and correct predictions is summarized with count values. The system scores perfectly when all the classified rates locate at the diagonal position of the matrix. If any data is misclassified, the values have deviated from diagonal cells.[24,25] Finally, the performance of the model was calculated using following performance measurement equations:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \tag{6}$$

$$\text{Sensitivity} = TP/(TP+FN) \tag{7}$$

$$\text{Specificity} = TN/(TN+FP) \tag{8}$$

$$\text{Precision} = TP/(TP+FP) \tag{9}$$

RESULTS AND DISCUSSION

All the analysis was done in MATLAB 2016 software with Intel Core i5, 2.4 GHz, and 6 GB RAM. The analysis of image pre-processing outcomes is shown in Figures 3 and 4.

Eight neighbor points ($P = 8$) with central radius (R=1) are applied with uniform patterns; hence, the feature vector length reduces from 256 gray scale to 59 binary patterns. Among 59 LBP extracted features, only 20 features were remained by BPSO feature selection which is statistically significant in medical imaging and other irrelevant features were canceled from the feature vector. The best solution achieved by the objective function in BPSO is 0.87 [Figure 5]. In Table 2, the performance measurements comparison is represented applying different functions (Linear, Polynomial, Gaussian) for SVM and (Fine, Cubic, Weighted) KNN. For passed features dataset, the SVM (Gaussian) classifier achieves the best accuracy 94.6%, Sensitivity 96.2%, Specificity 93%, and Precision 92.7%. The performance measurements are calculated by applying eq. (6) to eq. (9). The confusion matrix of the proposed method is tabulated in Table 3.

In Figure 6, the performance measurements comparison is represented applying different functions (Linear, Polynomial, and Gaussian) for SVM and (Fine, Cubic, Weighted) KNN. The passed features dataset, the SVM (Gaussian) classifier achieves the best accuracy 94.6%, Sensitivity 96.2%, Specificity 93%, and Precision 92.7%. The performance measurements are calculated by applying eq. (6) to eq. (9). The confusion matrix of the proposed method is tabulated in Table 2 below.

Figure 6 represents the bar graph to depict the comparison between the performance parameters of average specificity,

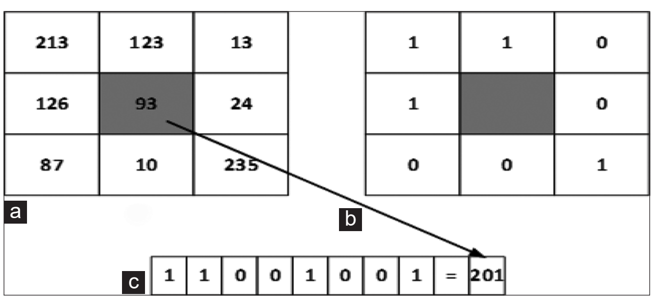

Figure 2: Local binary pattern process over gray scale pixels with Radius (R=1) and 8 neighbor pixels (a) observed pixel of gray scale image (b) binary encode between the center pixel and its neighbors (c) LBP string value and converted from binary to decimal

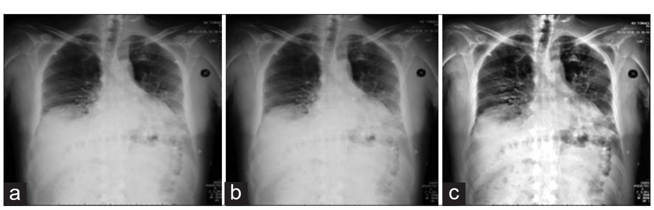

Figure 3: Results of affecting pre-processing filters on COVID-19 CXR images (a) original resized image (b) image denoising using Gaussian filter (c) image enhancement using Adaptive Histogram Equalization

Table 1: Binary particle swarm optimization algorithmic parameters for searching the best solution by swarms

Algorithm	Parameter name	Value
BPSO	Up (Upper band)	10
BPSO	Lb (Lower band)	−10
BPSO	Cognitive component (C1)	2
BPSO	Social Component (C2)	2
BPSO	Maximum inertia (Wmax)	0.9
BPSO	Minimum inertia (Wmin)	0.4
BPSO	Population Size	10
BPSO	Iteration	100

Table 2: Confusion matrix representation Gaussian support vector machine classification after ten-fold cross-validation

		Actual case	
Predicted case	51	4	55
	2	53	55
	53	47	110

	Actual class		
Predicted class	51	4	55
	2	53	55
	53	47	110

Table 3: Summarizes of classification results after calculation average accuracy, sensitivity, specificity, and precision from the confusion matrices

Classifiers	Type	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
KNN	Fine	90.9	86.9	95.9	96.4
	Cubic	84.5	83.9	93.2	94.5
	Weighted	93.6	92.9	94.4	94.5
SVM	Linear	91.8	77.9	82.4	76.4
	Polynomial	92.7	94.3	82.4	94.5
	Gaussian	94.6	96.2	93	92.7

Figure 4: Results of affecting pre-processing filters on normal CXR images (a) original resized image (b) image denoising using Gaussian filter (c) image enhancement using adaptive histogram equalization

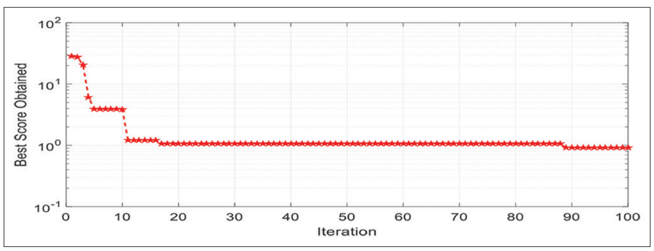

Figure 5: Best score achieved by objective function versus iteration for binary particle swarm optimization

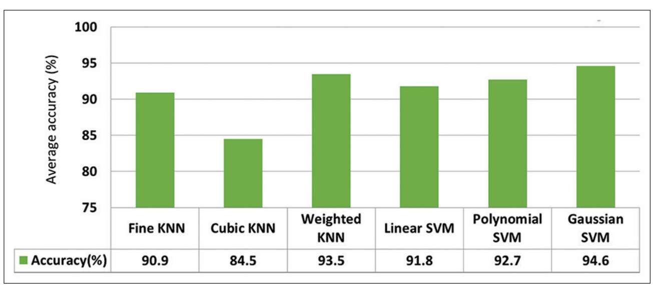

Figure 6: Average accuracy taken by used classifiers

sensitivity, and precision values achieved by Weighted KNN and Gaussian SVM.

For sake of a fair comparison between proposed work and the work presented in Baity[26] this research did not use any feature selection algorithm whereas in the proposed work an intelligent feature selection strategy has been used which is based on a practical Swarm optimization (bio-inspired technique), which is served as effective tool to filter out all irrelevant features before classification stage. In addition, two this work proposed two types of classifiers (SVM and K-NN), whereas in Baity[26] only one classifier (K-NN) had been suggested. Eventually, the maximum accuracy attained is 94.6%for COVID -19 virus detection which outperform the system best recognition accuracy attained in Baity[26] which was only 93.9%.

CONCLUSION

Since a few days after spreading this outbreak, many CAD methods for diagnosing COVID-19 have been proposed. There is no doubt that early detection by imaging machines can survive the current rate of contagious patients. Till now, many attempts are under progress to improve the performance measurements in data imaging for helping the doctors to diagnose that disease precisely. In this study for automatically classify 110 frontal CXR images according to normal, and COVID-19 stages, LBP was used to extract the texture features from the images after pre-processing step. The clinically significant features are selected by BPSO. Finally, the proposed method is evaluated by performing KNN and SVM classifiers. The results show that the system robustness in COVID-19 detection. Our experimental results show that the proposed method can be used as a second purpose to help the radiologists for classifying the frontal CXR images either positive or negative COVID-19.

REFERENCES

1. M. Y. Ng, E. Y. P. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M. S. Lui, C. S. Y. Lo, B. Leung, P. L. Khong, C. K. M. Hui, K. Y. Yuen, and M. D. Kuo. Imaging profile of the COVID-19 infection: Radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.
2. M. Thomas, G. Ksiazek, D, Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, and C. Urbani. *Respiratory and Enteric Virus Branch*. United States: National Center for Immunization and Respiratory Diseases, 2003.
3. M. Chan-Yeung, and R. Xu. SARS: Epidemiology. *Respirology*, vol. 8, pp. S9-S14, 2003.
4. A. Zumla, D. S. Hui, and S. Perlman. Middle East respiratory syndrome. *The Lancet*, vol. 386, no. 9997, pp. 995-1007, 2015.
5. A. J. Rodriguez-Morales, J. A. Cardona-Ospina, E. Gutiérrez-Ocampo, R. Villamizar-Peña, Y. Holguin-Rivera, J. P. Escalera-Antezana, L. E. Alvarado-Arnez, D. K. Bonilla-Aldana, C. Franco-Paredes, A. F. Henao-Martinez, A. Paniz-Mondolfi, G. J. Lagos-Grisales, E. Ramírez-Vallejo, J. A. Suárez, L. I. Zambrano, W. E. Villamil-Gómez, G. J. Balbin-Ramon, A. A. Rabaan, H. Harapan, K. Dhama, H. Nishiura, H. Kataoka, T. Ahmad, and R. Sah. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Medicine and Infectious Disease*, vol. 34, p. 101623, 2020.
6. H. Y. F. Wong, H. Y. S. Lam, A. H. T. Fong, S. T. Leung, T. W. Y. Chin, C. S. Y. Lo, M. M. S. Lui, J. C. Y. Lee, K. W. H. Chiu, T. W. H. Chung,

E. Y. P. Lee, E. Y. F. Wan, I. F. N. Hung, T. P. W. Lam, M. D. Kuo, and M. Y. Ng. Frequency and distribution of chest radiographic findings in COVID-19 positive patients. Radiology, vol. 296, No. 2, pp. 72-78, 2020.

7. L. Fan, D. Li, H. Xue, L. Zhang, Z. Liu, B. Zhang, L. Zhang, W. Yang, B. Xie, X. Duan, X. Hu, K. Cheng, L. Peng, N. Yu, L. Song, H. Chen, X. Sui, N. Zheng, S. Liu, and Z. Jin. Progress and prospect on imaging diagnosis of COVID-19. *Chinese Journal of Academic Radiology*, vol. 3, no. 1, pp. 4-13, 2020.
8. 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, and T. I. Islam. Can AI help in screening viral and COVID-19 pneumonia? *arXiv*, vol. 8, pp. 132665–132676, 2020.
9. A. Abbas, M. M. Abdelsamea, and M. M. Gaber. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Applied Intelligence*, vol. 51, no. 2, pp. 854-864, 2021.
10. K. Purohit, A. Kesarwani, D. R. Kisku, and M. Dalui. COVID-19 detection on chest X-Ray and CT Scan images using multi-image augmented deep learning model. *BioRxiv*. vol. 1, p. 205567, 2020.
11. A. Narin, C. Kaya, and Z. Pamuk. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, vol. 24, pp. 1207-1220, 2021.
12. K. El Asnaoui, and Y. Chawki. Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3615-3626, 2020.
13. S. H. Wady, R. Z. Yousif, and H. R. Hasan. A novel intelligent system for brain tumor diagnosis based on a composite neutrosophic-slantlet transform domain for statistical texture feature extraction. *BioMed Research International*, vol. 2020, p. 8125392, 2020.
14. B. N. S. Mohammed, and R. Z. Yousif. Intelligent system for screening diabetic retinopathy by using neutrosophic and statistical fundus image features. *Zanco Journal of Pure and Applied Sciences*, vol. 31, no. 6, pp. 30-39, 2019.
15. A. Malhotra, A. Sankaran, A. Mittal, M. Vatsa, and R. Singh. Fingerphoto authentication using smartphone camera captured under varying environmental conditions. In: *Human Recognition in Unconstrained Environments: Using Computer Vision, Pattern Recognition and Machine Learning Methods for Biometrics*. Netherlands: Elsevier, pp. 119-144, 2017.
16. Y. Cai, G. Xu, A. Li, and X. Wang. A novel improved local binary pattern and its application to the fault diagnosis of diesel engine. *Shock and Vibration*, vol. 2020, p. 9830162, 2020.
17. M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro, and C. S. Haley. Application of high-dimensional feature selection: Evaluation for genomic prediction in man. *Scientific Reports*, vol. 5, p. 151, 2015.
18. S. W. Lin, K. C. Ying, S. C. Chen, and Z. J. Lee. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, vol. 35, no. 4, pp. 1817-1824, 2008.
19. J. Too, A. R. Abdullah, and N. M. Saad. A new co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection. *Informatics*, vol. 6, no. 2, p. 21, 2019.
20. V. Kothari, J. Anuradha, S. Shah, and P. Mittal. A survey on particle swarm optimization in feature selection. *In Communications in Computer and Information Science*, vol. 270, pp. 192-201, 2012.
21. L. R. Fleah and S. A. Al-Aubi. A face recognition system based on principal component analysis-wavelet and support vector machines. *Cihan University-Erbil Scientific Journal*, vol. 3, no. 2, pp. 14-20, 2019.
22. R. Nisbet, G. Miner, and K. Yale. *Handbook of Statistical Analysis and Data Mining Applications*. Cambridge, Massachusetts: Academic Press, 2017.
23. S. Jang, Y. E. Jang, Y. J. Kim, and H. Yu. Input initialization for inversion of neural networks using k-nearest neighbor approach. *Information Science*, vol. 519, pp. 229-242, 2020.
24. E. El-Din 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*, vol. 2003, p. 11055, 2020.
25. D. R. J. Ramteke, and K. Y. Monali. Automatic medical image classification and abnormality detection using K-nearest neighbour. *International Journal of Advanced Computer Research*, vol. 2, no 4, p. 6, 2012.
26. S. Baity. COVID-19 detection for chest X-ray images using local binary pattern. *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 78-81, 2020.