Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd

ISSN,e-ISSN: 2548-1304
Volume:IDAP-2022, pp:120-129, 2022
https://doi.org/ 10.53070/bbd.1172671
Research Paper

A Novel Covid-19 Detection System Based on PSO and Hybrid Feature Using Support Vector Machines

Mehmet F. Ozdemir[*1,2], Davut Hanbay[1]

[1]Dept. of Computer Engineering, Inonu University, Malatya, Turkiye

[2]HAVELSAN Inc. Ankara, Turkiye,

(mfatih.ozdemir@inonu.edu.tr, davut.hanbay@inonu.edu.tr)

Received:Sep.08,2022 Accepted:Sep.16,2022 Published:Oct.10,2022

Abstract— The world first met the coronavirus (COVID-19) in Wuhan, China in December 2019. It has continued to increase its influence from the first encounter until today. The detection of this virus, which has caused the death of many, is of great importance today. There are many approaches to the detection of this disease. One of the most effective of these approaches is the detection of COVID-19 disease using chest X-Ray images. In this paper, an intelligent system was proposed to classify normal, pneumonia patients and COVID-19 patients using chest X-Ray images. The proposed system was composed of four stage. At first, all images in the dataset were pre-processed. Then for the feature extraction uniform Local Binary Pattern (LBP) and DenseNet201 deep learning models were used. Particle swarm optimization (PSO) algorithm was used to select effective features. The determined effective features were classified by support vector machine (SVM). Accuracy and AUC parameters were used as performance criteria. Evaluated accuracy and AUC values were 99.9%, 1.00, respectively. The dataset and proposed model codes are made publicly available at: https://github.com/mfatiho/covid-detection-chest-xray

Keywords: *Covid-19, Feature Selection, LBP, DenseNet, SVM*

1. Introduction

Covid-19 was first reported at the World Health Organization (WHO) China Office in December 2019. Chinese authorities first identified a new type of coronavirus (novel coronavirus) on January 7, 2020 and WHO declared COVID-19 outbreak a global pandemic in March 2020 (Cucinotta & Vanelli, 2020). After the covid-19 was declared as global pandemic, it was determined that it showed more than one symptom. According to studies, the most common symptoms were fever, cough, fatigue (Guan et al., 2020; C. Huang et al., 2020).

Rapid diagnosis has a very important place in today's world where cases are increasing rapidly. One of the most commonly used auxiliary methods for this is the analysis of chest x-ray images. In addition to this method, chest CT scans and RT-PCR tests are used in the diagnosis of the disease.

On the other hand, the use of medical images to diagnose diseases has increased. Different computer vision and image processing tools can be used to diagnose diseases. One of the main advantages of using this type of system is that detects fast and accurate, as well as reducing the need for manpower. Medical image processing algorithms can give clues about a possible disease and can be used as a diagnosis. In COVID-19 disease, X-ray images allow to examine how the virus affects the lungs and are used as an auxiliary step in diagnosis.

It is classified by many different techniques using chest x-ray images. Machine learning and its sub-study deep learning are among the most used. While deep learning algorithms require a lot of data to train the network, on the other hand, machine learning needs less data. Covid-19 chest x-ray images are still not easy to find today. Therefore, it would be a better choice to use machine learning for classification. Many studies are available in this area using deep learning or machine learning.

In (Göreke et al., 2021), a new set of features was obtained based on laboratory findings, taking the ethnic and genetic differences in human blood into account. Later, a new hybrid classification application and COVID-19 detection were made to deep learning within this feature group. In binary classification (NON COVID-19, COVID-19), 94.98% precision, 94.98% recall, 94.95% accuracy, 94.98% F1 score and 100% AUC performance data were achieved.

In (Wu et al., 2021), chest X-ray was used for 7 different scenarios according for healthy, bacterial, viral and COVID-19 classes. To build proposed architecture, a combination of Generative Adversarial Networks (GANs), transfer learning and LSTM networks were used. In addition, they used GANs to increase number of samples in their dataset so the number of 317 Covid-19 images was increased to 2842. In the studied scenarios, the highest accuracy in two classifications was found to be 99.5%, the highest accuracy in the triple classification was 98.1%, and the quadruple classification was 91.8%.

In (Karakanis & Leontidis, 2021), two deep learning models were proposed. The first of these was binary model and the other model was for multiple classification. They achieved 100% sensitivity, 98.3% specificity and 98.7% accuracy for binary classification and they achieved 99.3% sensitivity, 98.1% specificity and 98.3% accuracy for multiple classification.

In (Ismael & Şengür, 2021), Covid-19 and healthy images were classified using three different deep learning approaches. The accuracy of each was evaluated separately. As a result, it was stated that using SVM classifier with deep features gives better result.

In (Varela-Santos & Melin, 2021), they proposed to classify data sets of medical images from COVID-19 patients and some other related diseases affecting the lungs. They developed a system that can automatically detect COVID-19 disease based on chest x-rays and computed tomography images of the lungs. As a result of four different experiments, they achieved 100% accuracy, 1.0 AUC in binary classification (COVID-19, No Finding) and 98.56% accuracy, 1.0 AUC in multiple classification (COVID-19, No Finding, Pneumonia)

In this paper, we conducted three studies for classification of chest x-ray images. One of them was for classifying Normal and COVID-19, the second one was for classifying Pneumonia and COVID-19, the last one was for classifying Normal, Pneumonia and COVID-19. The used chest x-ray images were collected from different datasets. The achieved accuracy and AUC values for first model were 99.9%, 1.0, respectively. The achieved accuracy and AUC values for second model were 99.9%, 1.0, respectively. The achieved accuracy and AUC values for third model were 99.9%, 1.0, respectively.

The organization of the paper is as follows, the methods and algorithms used in the study were explained briefly in material and methods section. The details of the application were given in third section, results and conclusions were given in section four.

2. Material and Methods

In this section we introduced briefly the basics of used methods

2.1. Local Binary Pattern

LBP method (Ojala et al., 1996), which is gray-scale invariant, is one of the best visual descriptors. The value in the center of n x n neighborhood values is used as the threshold. After thresholding is applied to the values of the neighborhood pixels, the weights calculated with power of two are multiplied by the corresponding pixels. The LBP value is calculated by the Eq.1

$$LBP_{P,R} = \sum_{n=0}^{P-1} S * 2^n \, S = \begin{cases} 1 & (T_i - T_c) \geq 0 \\ 0 & (T_i - T_c) < 0 \end{cases} \qquad (1)$$

where T_c represents the gray value of the center pixel of local neighborhood and T_i represents the gray values of P pixels on circle of radius R forming a circular symmetrical set of neighbors. A sample LBP calculation schema is shown in Fig.1

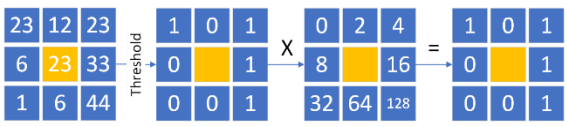

Fig.1. LBP calculation schema

After the basic LBP method was created, many new approaches were built on it. Uniform patterns in LBP used by (Ojala et al., 2002) are just one of them. When a pattern is placed circularly, if the 0-1 change consists of two transactions, it is called uniform. For example, 00011100, 11100000, 10000111 uniform, 01010101, 10001011, 01111011 nonuniform.

2.2. DenseNet

Deep learning methods have started to grow and develop rapidly with the developing hardware technologies. With the development of deep learning, many convolutional neural networks have been brought to the literature and used effectively. One of the most important of these is Dense Convolutional Network (DenseNet).

DenseNet (G. Huang et al., 2017) model has a few numbers of parameters because it does not need to relearn unnecessary feature maps. It tries to cope with the vanishing-gradient problem more strongly. It allows each layer to have strong control as it has direct access to gradients from loss function and original input signal. Feature propagation is strong and encourages feature reuse. Communication between layers is strong and all layers are directly interconnected. The features are added to the next layer and combined. The basic DenseNet arthitecture is shown in Fig.2

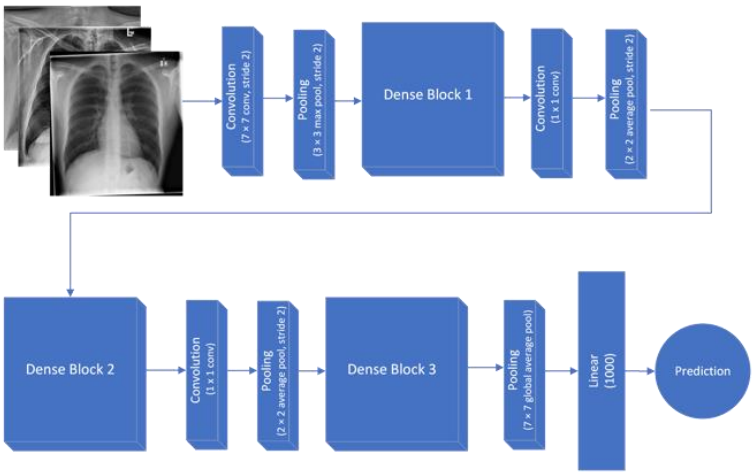

Fig. 2. DenseNet Architecture

2.3. Particle Swarm Optimizer (PSO)

The purpose of the particle swarm optimizer (PSO) is to mathematically model the movements of a flock of birds without crashing. The optimizer was developed originally by (Clerc, 2010). PSO is effective and easy to manage and implement parameters. Compared to other heuristic approaches, computational efficiency is quite successful. The velocity and position of each particle is refreshed using Eq. 2

$$v_{(t+1)}^{i} = v_{(t)}^{i} + c_1 r_1 \left(pbest_{(t)}^{i} - x_{(t)}^{i}\right) + c_2 r_2 \left(gbest - x_{(t)}^{i}\right) \qquad (2)$$

where $x_{(t)}^{i}$ represents ith particle at current iteration (t), $v_{(t)}^{i}$ represents velocity of the ith particle at current iteration (t), c_1 corresponds to cognition learning parameter, c_2 corresponds to social learning parameter, r_1 and r_2 are random numbers in the range 0 to 1, $pbest_{(t)}^{i}$ represents local/personal best position for ith particle at current iteration (t), $gbest$ represents global best position for all particles. $v_{(t+1)}^{i}$ represents velocity of the ith particle at next iteration $(t+1)$.

All particles are repositioned according to fitness function. Each particle compares the fitness value in the previous best position with the fitness value in the current position to find the new best position. $c_1 r_1 \left(pbest_{(t)}^{i} - x_{(t)}^{i}\right)$ in the

formula is used for this. In addition, $c_2 r_2 \left(gbest - x_{(t)}^{i} \right)$ in the formula is used to bring the particles closer to the global best position of all particles.

2.4. Support Vector Machine (SVM)

The SVM is a supervised learning approach which is developed by (Cortes & Vapnik, 1995) and his colleagues at AT&T Bell Labs in 1995. It can be applied to both regression and classification tasks. The SVM approach has adopted the principle of Structural Risk minimization. The output of the classifier can be evaluated by Eq.3.

$$y(x) = sign[\sum_{k=1}^{N} a_k \, y_k x_k^T x + b] \qquad (3)$$

The linear SVM classifier was extended to inseparable case by (Cortes & Vapnik, 1995). It is done by taking the additional slack variable in the problem formulation given in Eq.4.

$$y_k[w^T x_k + b] \geq 1 - \, \xi_k = \pi r^2, k = 1, \ldots, N \qquad (4)$$

SVM has also been used for nonlinear and linear function estimation, among other uses. Details are available in the literature. (Hanbay, 2009)

3. Application

The proposed method was implemented using a system that have 16 GB RAM with an intel i7 (9750H) processor. All program code were coded in Matlab R2020a. To evaluate the model performance 5-fold cross validation method was used. Accuracy and AUC parameters were used as performance criteria. The block diagram of the proposed method is shown in Fig.3.

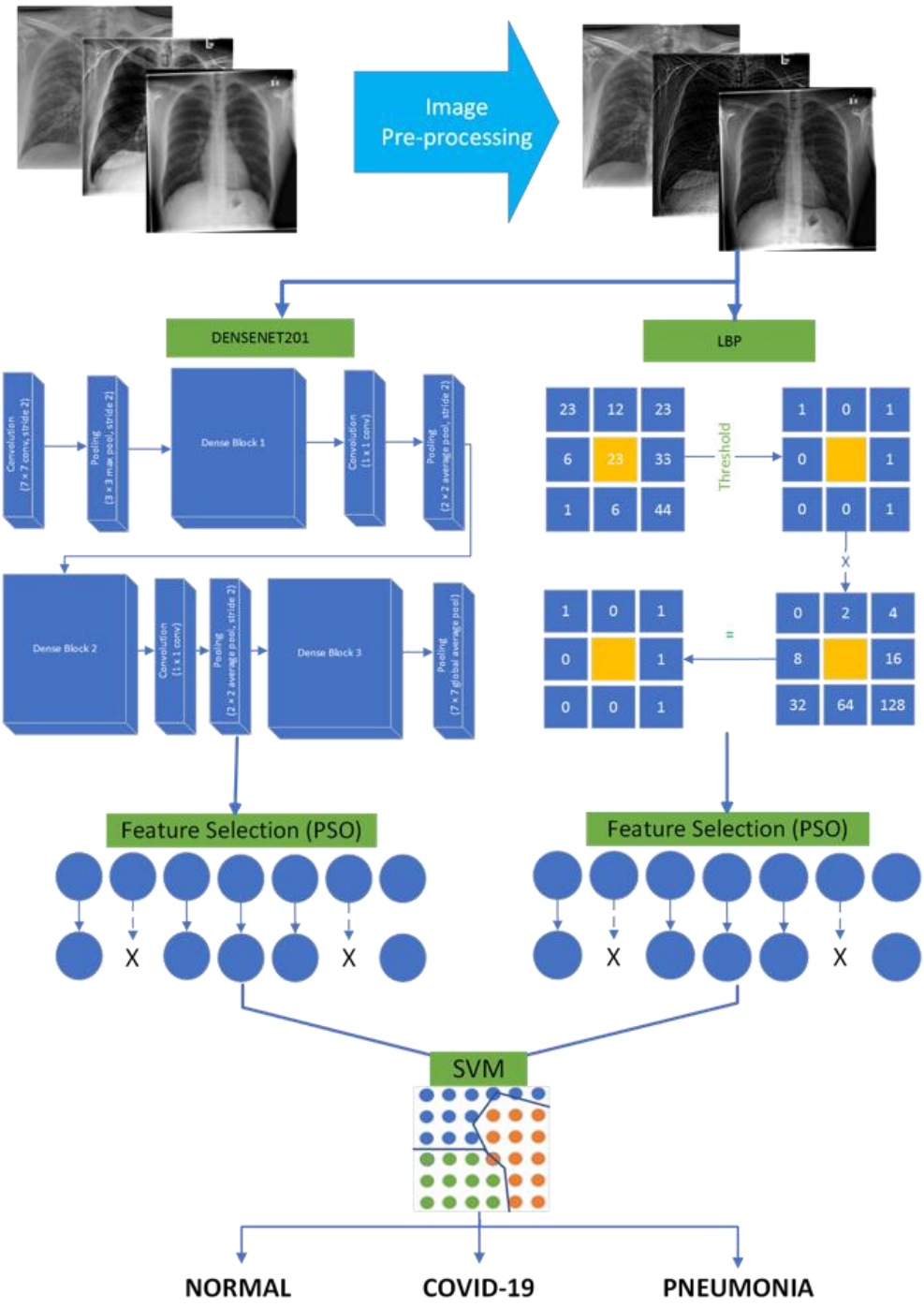

Fig. 3. The block diagram of the proposed method.

3.1. Dataset

A new dataset was created to be used in this study using three different datasets. The first dataset is provided by (Cohen et al., 2020). This dataset contains several frontal chest x-rays. It is one of the largest publicly available dataset for COVID-19. 468 front chest x-rays images are taken from dataset one. Second dataset is provided by a group of scientists at Stanford University. This dataset called as CheXpert. This dataset contains 224,316 chest radiograph

images of 65,240 patients. The dataset has been made public as a benchmark to evaluate the performance of chest radiography interpretation models. 4023 pneumonia images are taken from CheXpert dataset (Irvin et al., 2019). The third dataset is provided by a team of researchers from the University of Qatar, Doha, Qatar and Dhaka University, Bangladesh for COVID-19 positive cases. 10,192 normal chest x-ray image are taken from this dataset (*COVID-19 Radiography Database | Kaggle*, n.d.).The details of dataset used in this application are given in Table 1.

Table 1. The details of dataset used in this application.

Dataset	COVID-19	Normal	Pneumonia
COVID-19 (Cohen et al., 2020)	468	0	0
CheXpert (Irvin et al., 2019)	0	0	4023
COVID-19 Radiography Database (*COVID-19 Radiography Database | Kaggle*, n.d.)	0	10192	0
Total	468	10192	4023

3.2. Image Pre-processing

Before applying all images to the proposed model, they were pre-processed as shown in Fig. 4. Atmospheric blur is reduced and all images resized to 299x299.

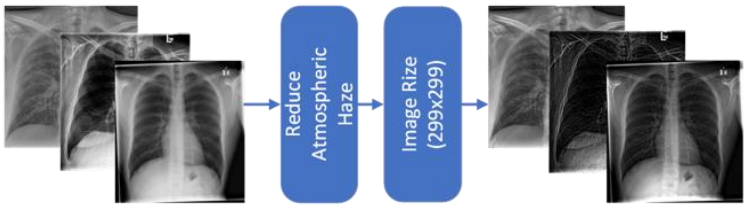

Fig. 4. Image pre-processing

3.3. Feature Extraction

After image pre-processing, a few of deep learning approaches and traditional methods were tested empirically to select the best feature. LBP, Histogram of Oriented Gradients (HOG), Resnet18, Mobilnetv2, Densenet201 were experimentally tested. The best performance was obtained with the combination of LBP and Densenet201. The block diagram of the feature extraction step is shown in Fig.5.

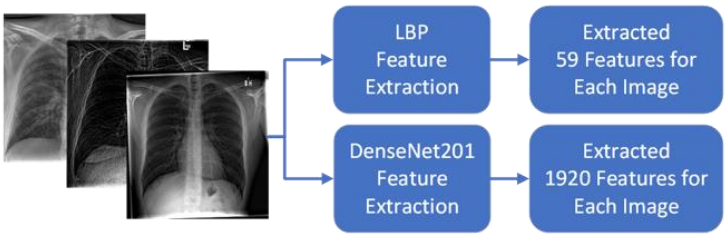

Fig. 5. The block diagram of feature extraction

59 features were extracted by LBP and 1920 features were extracted by Densenet201 deep learning model from each image. In the Densenet201 deep learning model, features were extracted in the average pooling layer before the fully connected layer.

3.4. Feature Selection

Feature selection is the stage of selecting the most valuable features for classification. Particle swarm optimization, one of the most successful optimization algorithms, so was chosen for feature selection. The process of selecting the most valuable features is shown in Fig. 6.

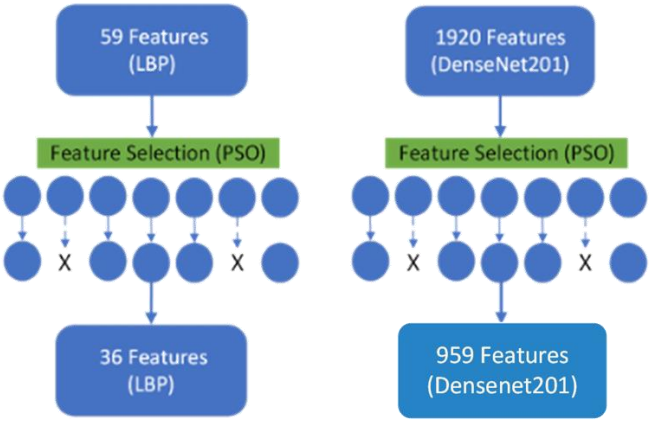

Fig. 6. Feature selection process

30 particles were used when selecting features in PSO. In Eq.2 we use to update the velocities and positions of the particles, the constants c_1, c_2 are 2 and r_1, r_2 in the equation we use are random values between 0 and 1.

As a result, the most valuable 36 features were selected out of 59 features in LBP, and 956 features were the most valuable out of 1920 features in Densenet201.

3.5. Classification

After feature selection, selected feature was applied to the SVM model. Three different classification processes were made with SVM.

At first, COVID-19 and Normal images were classified. Evaluated accuracy and AUC values are 99.9% and 1.0, respectively. This classification confusion matrix and ROC curve shown in Fig. 7.

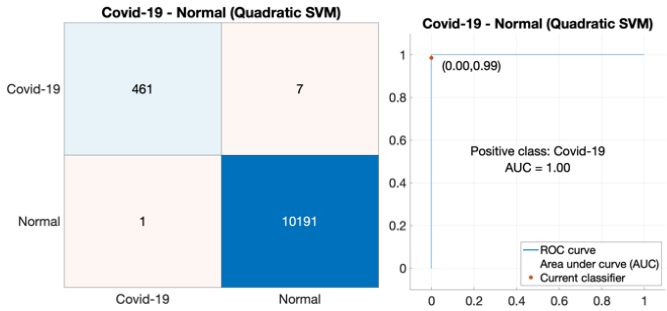

Fig. 7. Confusion matrix and ROC curve of model-1

Secondly, Covid-19 and Pneumonia images were classified. Evaluated accuracy and AUC values are 99.9% and 1.0, respectively. This classification confusion matrix and ROC curve shown in Fig. 8.

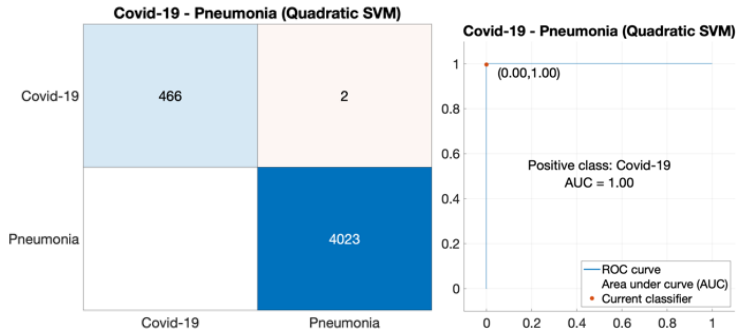

Fig. 8. Confusion matrix and ROC curve of model-2

Thirdly, Covid-19, Pneumonia and Normal images were classified. Evaluated accuracy and AUC values are 99.9% and 1.0, respectively. This classication confusion matrix and ROC curve shown in Fig. 9.

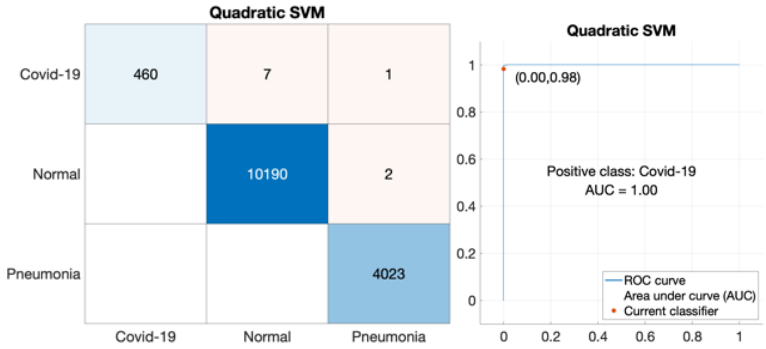

Fig. 9. Confusion matrix and ROC curve of model-3

4. Conclusion

Many COVID-19 studies have been carried out to this date and several deep learning, feature extraction and classification methods were used in most of these studies. In current study, three different types of chest x-ray images were classified with the developed system. First, DenseNet201 and LBP were used for feature extraction from images. Afterwards, the most valuable features were selected using PSO. Finally, selected features were classified using SVM. As a result, 99.9% accuracy, 1.0 AUC score was obtained in COVID-19 and normal classification. In the COVID-19 and pneumonia classification, an accuracy of 99.9% and a score of 1.0 AUC were obtained. In COVID-19, pneumonia and normal classification, 99.9% accuracy and 1.0 AUC score were obtained. With the proposed method, a very successful result has been achieved in COVID-19 classification. The classification success of the current study can be used as a reference for different studies in the future.

Acknowledgments

This study was supported by Inonu University Scientific Research Projects Coordination Unit (BAP) with the project coded FYL-2021-2449.

Declarations

Ethics Approval The authors has confirmed that no ethical approval is required

Confict of Interest The authors declare that they have no confict of interest.

References

Clerc, M. (2010). Particle Swarm Optimization. *Particle Swarm Optimization*, 1942–1948. https://doi.org/10.1002/9780470612163

Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). *COVID-19 Image Data Collection: Prospective Predictions Are the Future*. http://arxiv.org/abs/2006.11988

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

COVID-19 Radiography Database | Kaggle. (n.d.). Retrieved April 14, 2021, from https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomedica*, *91*(1), 157–160. https://doi.org/10.23750/abm.v91i1.9397

Göreke, V., Sarı, V., & Kockanat, S. (2021). A novel classifier architecture based on deep neural network for COVID-19 detection using laboratory findings. *Applied Soft Computing*, *106*, 107329. https://doi.org/10.1016/j.asoc.2021.107329

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., … Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, *382*(18), 1708–1720. https://doi.org/10.1056/nejmoa2002032

Hanbay, D. (2009). An expert system based on least square support vector machines for diagnosis of the valvular heart disease. *Expert Systems with Applications*, *36*(3 PART 1), 4232–4238. https://doi.org/10.1016/j.eswa.2008.04.010

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019). *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*. http://arxiv.org/abs/1901.07031

Ismael, A. M., & Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications*, *164*(March 2020). https://doi.org/10.1016/j.eswa.2020.114054

Karakanis, S., & Leontidis, G. (2021). Lightweight deep learning models for detecting COVID-19 from chest X-ray images. *Computers in Biology and Medicine*, *130*(November 2020), 104181. https://doi.org/10.1016/j.compbiomed.2020.104181

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, *29*(1), 51–59. https://doi.org/10.1016/0031-3203(95)00067-4

Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant

texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623

Varela-Santos, S., & Melin, P. (2021). A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks. *Information Sciences*, *545*, 403–414. https://doi.org/10.1016/j.ins.2020.09.041

Wu, X., Chen, C., Zhong, M., Wang, J., & Shi, J. (2021). COVID-AL: The diagnosis of COVID-19 with deep active learning. *Medical Image Analysis*, *68*, 101913. https://doi.org/10.1016/j.media.2020.101913