Covid-19 detection from radiographs by feature-reinforced ensemble learning

Abdullah Elen

Department of Software Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul University, Bandirma, Balikesir, Turkey

Correspondence
Abdullah Elen, Department of Software Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul University, 10200 Bandirma, Balikesir, Turkey. Email: aelen@bandirma.edu.tr

Summary
The coronavirus (Covid-19) epidemic continues to have a negative influence on the global population's well-being and health. Scientists in many fields around the world are working non-stop to find a solution to the prevention of this epidemic. In the field of computer science, this struggle is supported by studies on especially the analysis of X-ray and CT images with artificial intelligence. In this study, two different ensemble learning models, including deep learning and a combination of machine learning methods, are presented for the detection of SARS-CoV-2 infection from X-ray images. The main purpose of this study is to increase the classification ability of Residual Convolutional Neural Network (ResCNN), which is used as a deep learning method, with the assist of machine learning algorithms and extracted features from images. The proposed models were validated on a total of 5228 chest X-ray images categorized as Normal, Pneumonia, and Covid-19. The images in the dataset were sized in four different ways, 32 × 32, 64 × 64, 128 × 128, and 256 × 256, in order to analyze the validity of the proposed models in more detail. These four datasets were partitioned with the 10-fold cross-validation technique and converted into a total of 40 training and test data. Both proposed models use features derived from the ResCNN as the basis and test a certain number of machine learning algorithms with a majority voting technique by dividing them into subsets. In the architecture of the second model, it combines the features extracted from the Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) methods in addition to the features obtained from the ResCNN. It has been seen that the classification ability of both proposed models is better than the ResCNN in the experiments. In particular, the second model gives a similar classification score even though it is tested with images four-times smaller (e.g., 32 × 32 vs. 128 × 128) than those used in the ResCNN. This shows that the model can give ideal results with lower computational cost.

KEYWORDS
convolutional neural network, Covid-19, histogram-oriented gradients, local binary patterns, machine learning, X-ray images

Abbreviations: CNN, convolutional neural network; DT, Decision Tree; FCL, fully connected layer; HOG, Histogram of Oriented Gradients; k-NN, k-Nearest Neighbor; LBP, Local Binary Patterns; LDA, Linear Discriminant Analysis; MV, majority voting; NB, Naïve Bayes; NCA, Neighborhood Component Analysis; PM1, PM2, proposed model #1 and #2; ResCNN, Residual Convolutional Neural Network; SGDM, Stochastic Gradient Descent with Momentum; SVM, Support Vector Machine.
1  INTRODUCTION

The novel coronavirus disease (Covid-19) is a virus that was initially found on January 13, 2020, as a result of a study undertaken in a group of individuals who acquired respiratory symptoms (fever, cough, and shortness of breath) in late December 2019 in Wuhan, China.1 The pandemic was first discovered in people who worked in the seafood and livestock markets in this area.2,3 Later, it expanded from person to person and to other cities in Hubei province, particularly Wuhan, as well as other provinces of China and other nations around the world.4 Coronaviruses are a broad group of viruses that can infect both animals and people. Several coronaviruses have been linked to respiratory infections in humans, ranging from the common cold to more serious illnesses including Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS).5 The SAR-CoV-2 virus causes novel coronavirus illness. Although it has been reported that there may be asymptomatic cases in the symptoms of the novel coronavirus disease, the rate of these is not known.6 Fever, cough, and shortness of breath are the most prevalent symptoms.7 Pneumonia, severe respiratory failure, renal failure, and death may occur in extreme cases. Covid-19 virus is transmitted by coughing and sneezing of sick individuals and inhaling droplets scattered around the environment. The virus can be transmitted to the eyes, face, nose, and mouth without washing the hands after contacting surfaces infected with respiratory particles. Contact with eyes, nose or mouth with dirty hands is risky.8 According to the information available thus far on Covid-19 infection, some people are at a higher risk of becoming ill and having severe symptoms. Accordingly, 80% of the cases have mild disease and 20% are treated in hospital conditions.9 People with medical conditions such as heart disease, hypertension, diabetes, cancer, and chronic respiratory disease; those over the age of 60, and healthcare workers are most affected by the disease.10

Machine learning algorithms can be incorporated to build an automatic computer-aided diagnostic (CAD) system to struggle the Covid-19 pandemic.11 A significant number of clinical and radiological research have been published in this field, reporting various radiographic imaging results as well as the epidemiology of Covid-19.12 Many deep-learning models, such as convolutional neural networks (CNN),13–16 transfer learning models,17,18 generative adversarial networks (GAN),19,20 and others, have also been used to automatically analyze diseases diagnosed by radiological methods. Covid-19 is identified utilizing chest X-rays in studies with multiple or binary classifications. Raw image data is used in certain research, whereas feature extraction techniques is used in others. The number of samples in the datasets employed in investigations varies as well. Research in this area has shown that CNNs are the most popular classifiers for corona virus detection. Ai et al.21 to evaluate performance of chest computed tomography (CT) with use of the reverse-transcription polymerase chain reaction (RT-PCR) in the diagnosis of Covid-19, 1014 patients who underwent both chest CT and RT-PCR tests in Wuhan, China were included in their study. As a result of statistical analyzes with SPSS, they reported that chest CT has a high sensitivity for the diagnosis of Covid-19 and therefore can be considered as the primary tool for corona virus detection in epidemic areas.

Some of these studies are as follows: Waheed et al.22 developed a method based on the Auxiliary Classifier Generative Adversarial Network (ACGAN), called CovidGAN, to produce synthetic chest X-ray images for the dataset containing a limited number of X-ray images used in their studies and to increase the success of CNN in detecting Covid-19. In their experimental studies, they reported that they achieved 85% classification accuracy with the use of CNN alone, and this rate increased to 95% when they added synthetic images produced by CovidGAN. Ozturk et al.23 proposed two deep learning-based models to detect and classify Covid-19 cases from X-ray images. The first is the binary classifier, which labels “Covid” and “No-findings,” and the second is the multiclass classifier, which labels “Covid,” “No-findings,” and “Pneumonia.” As a result of their experimental studies, they reported that the proposed methods had an accuracy of 87.02% and 98.08%, respectively. Ismael and Şengür,24 proposed their experimental studies, they reported that the proposed methods had an accuracy of 99.1% with binary classification labeled as “Covid” and “Normal,” 94.2% with triple classification labeled as “Covid,” “Pneumonia,” and “Normal,” 91.2% with quadruple classification labeled as “Covid,” “Viral Pneumonia,” “Bacterial Pneumonia,” and “Normal.” Chandra et al.25 proposed a Covid screening system, called ACoS, which detects healthy, suspected, and corona virus-infected patients using radiomic texture descriptors extracted from X-ray images. The ACoS has a two-stage classification approach using a majority vote-based classifier set of five supervised learning methods. They used 2088 and 258 chest X-ray images for training-test and validation, respectively. According to the classification results obtained, they achieved an accuracy rate of 98% in the first phase and 91% in the second phase. Barstugan et al.26 to classify X-ray images for the diagnosis of corona virus using five different feature extraction methods that are Local Directional Patterns (LDP), Discrete Wavelet Transform (DWT), Grey-Level Size Zone Matrix (GLSZM), Grey-Level Run Length Matrix (GLRLM), and Grey-Level Cooccurrence Matrix (GLCM). These features were classified by SVM using 2-fold, 5-fold, and 10-fold cross validation techniques. They reported that the highest classification accuracy rate was 99.68% using the 10-fold cross-validation for features extracted from the GLSZM. Narin et al.27 used five different pre-trained CNN classifiers to detect Covid-19 cases using chest X-ray images. In their experimental study, they used 5-fold cross validation on three different datasets and reported the highest accuracy rate between 96% and 99.7% with ResNet50.
The contributions of this study are listed as follows:

1. A novel ensemble learning model is presented for diagnosing Covid-19 infection from chest X-ray images. The proposed model differs from traditional machine learning and deep learning methods in terms of its architecture.

2. The proposed study trains a certain number of machine learning algorithms using features extracted from Residual Convolutional Neural Network (ResCNN), Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). It separates and tests them into subsets and returns the subset with the highest classification accuracy based on majority voting as the solution.

3. By combining the power of deep learning and machine learning algorithms, it provides higher performance and is very practical to implement due to its automatic feature extraction.

4. Applicable to all classification problems involving images, as it does not need problem-specific parameters.

The rest of paper is organized as follows: The Section 2 contains technical details and sample images about the dataset used in the experimental studies. In Section 3, there are detailed explanations about the LBP and HOG, which are feature descriptor methods. In Section 4, proposed classification models for detection of Covid-19 infection from chest X-ray images are explained in detail. In addition, technical details about the ResCNN architecture used for comparison in experimental studies are given. In Section 5, the proposed models validated by experimental study on the dataset and classification performances of the ResCNN, PM1, and PM2 were compared. In addition, performance of the PM2 according to class labels is documented in detail. In the last section, this paper is summarized.

2 DATASET

The dataset is a collection of 5228 Covid-19 chest X-ray images (radiographies) available from Kaggle, a subsidiary of Google LLC. Specifically, the images are divided into three classes as shown in Table 1; the total number of normal cases is 1626, bacterial infection cases (Pneumonia) 1802 and viral infection cases (Covid-19) 1800 confirmed data. This dataset is publicly available on Kaggle. Figure 1 shows all three classes of radiographic X-rays.

All images in the dataset are 256 × 256 pixels and in PNG format. In this study, the original images were resized in three different sizes, 32 × 32, 64 × 64, and 128 × 128 pixels. Thus, four image datasets of different sizes were obtained, allowing for more detailed analysis of methods for the Covid-19 recognition.

| Class label | Abbreviation | Number of images |
|-------------|--------------|-----------------|
| Normal      | NOR          | 1626            |
| Pneumonia   | PNE          | 1802            |
| Covid-19    | COV          | 1800            |

Figure 1 shows all three classes of radiographic X-rays.
3 | FEATURE EXTRACTION

Feature extraction plays an important role in pattern recognition and classification.\textsuperscript{29} Appropriate and distinctive features can effectively present a variety of content in an image, while also providing a solid basis for ultimate classification. In this study, in addition to the features obtained from Residual CNN, features were extracted by applying LBP and HOG techniques to X-ray images in the dataset. Then, all these features were brought together and used in training and testing the proposed method.

3.1 | Local binary patterns

LBP is a type of visual descriptor (or texture operator) used in computer vision for classification.\textsuperscript{30,31} The LBP is calculated by thresholding intensity of circularly symmetric neighboring pixels with the intensity of the center pixel, in a local region of an image, as shown in Figure 2. In this respect, it can be seen as a unifying approach to the typically diverse statistical and structural concepts of texture analysis. In the calculation process, the neighbors with positive density differences are set to one and the others to zero. After this operation, the resulting binary (0 and 1) values are converted to decimal numbers. When the pixels surrounding the center are all black or all white, this image region is flat (i.e., there are no features). Consecutive groups of black or white pixels in the local region are considered "uniform" patterns that can be interpreted as vertices or edges. If the pixels alternate between black and white pixels, the pattern is considered "non-uniform."

The LBP has become a prominent technique in a variety of applications due to its computational simplicity and distinctive power. The most important property of the LBP in real-world applications can be said to be resistant to monotonic grayscale changes caused by illumination variations in images. In Equation (1), the notation \((P, R)\) is used for pixel neighborhoods that means \(P\) sampling points on a circle of radius of \(R\). Let \(I\) be a grayscale image, and \(I(x, y)\) to be a gray value at \((x, y)\) in \(I\). The LBP at \((x, y)\) is therefore defined as:

\[
\text{LBP}_{P,R}(x, y) = \sum_{i=0}^{P-1} \text{sign}(I(x_i, y_i) - I(x, y)) \cdot 2^i
\]  

where, \(I(x_i, y_i)\) is intensity of \(P\) equally spaced pixels in a circle with radius \(R\) centered on pixel \((x, y)\), and \((x_i, y_i)\) is the location of neighbors given by \(X\). If the neighbors are not in the center of the pixels, interpolation should be used to approximate their intensities. Equation (2) shows how to calculate the \(\text{LBP}_{P,R}\) value using \((P, R)\), as in Figure 3.

---

**FIGURE 2** Result samples of the LBP relative to the central pixel

**FIGURE 3** The LBP with different \((P, R)\): (4, 1), (8, 1)
\[
\text{sign}(x) = \begin{cases} 
1, & x \geq 0 \\
0, & \text{otherwise}
\end{cases}
\] (2)

Since LBP produces \(2^P\) different output values, the LBP value of each pixel in an image can be calculated and generate a \(2^P\)-bin histogram as the image descriptor.

The original LBP provides invariance to any monotonic transformation and scaling of the gray scale. Each rotation invariant LBP is given a unique ID, which is defined as stated in Equation (3), to achieve rotation invariance.

\[
\text{LBP}_{p,R}(x,y) = \min \left\{ \text{ROR} \left( \text{LBP}_{p,R}(x,y), s \right) \right\} \quad \text{for} \quad s = 0, 1, \ldots, P - 1
\] (3)

where, \(\text{ROR}(\text{LBP}_{p,R}(x,y), s)\) performs bitwise circular right shift.

### 3.2 Histogram of oriented gradients

The histogram of directed gradients (HOG) is a feature descriptor used in machine vision and image processing for object detection and image classification processes.\(^{32}\) The HOG identifier focuses on the structure or shape of an object.\(^{33}\) It is better than any edge descriptor as it uses magnitude as well as gradient angle to calculate features. Generates histograms for regions of the image using the gradient size and orientations. The purpose of this method is to express the image as local histograms. These histograms contain the number of repeats at certain offset values of the gradient orientations of the image. This simple but effective feature extraction method has gained intense interest in the literature due to its successful applications. Especially in CNN and computer vision, it has found a wide area of use on different problems.\(^{34}\) In the HOG method, the gradient values for each pixel in the image are first calculated using Equation (4).

\[
g_x(x,y) = I(x+1,y) - I(x,y) \\
g_y(x,y) = I(x,y+1) - I(x,y)
\] (4)

where, \(g_x\) and \(g_y\) are the horizontal and vertical gradient, respectively, and \(I(x,y)\) is the pixel value at position \((x,y)\) of the image. The gradient magnitude and direction are then calculated using Equation (5).

\[
\text{Magnitude}(\mu) = \sqrt{g_x^2 + g_y^2} \\
\text{Angle}(\theta) = \arctan \left( \frac{g_y}{g_x} \right)
\] (5)

After the gradient of each pixel is obtained, the gradient matrix is divided into \(k \times k\) cells and descriptive blocks \((C_{m,n})\) are created. By default, the value of \(k\) is usually taken as 8. In the next step, an \(n\)-bins histogram \((H)\) in the range of 0–180° is calculated for each block, as shown in Figure 4.

Here \(n\) is used to specify the angle range, which is usually 9. Thus, the step size becomes \(\Delta \theta = 180/n\), that is, 20°. The values to be given to the \(h\)th and \((h+1)\)th divisions, respectively, for each cell in a block are calculated by Equation (6). Here \(h = \lfloor \omega \rfloor\) and \(\omega = (\theta/\Delta \theta) - 0.5\).

\[
V_h = \mu \omega, \quad V_{h+1} = \mu (\omega - h)
\] (6)

Because a block contains 64 different values, calculations are made for all 64 magnitude and gradient values. When the histogram calculation for all blocks is finished, four blocks from the 9-bins histogram matrix are combined to form a new \(2 \times 2\) grid box \((B)\) as in Figure 5.

For all 4 cells in a grid box, combining all the 9-bin histograms for each block to form 36 feature vectors, yielding \(B_i = [H_1, H_2, \ldots, H_{36}]\). In the last step, the \(B_i\) values for each grid box are normalized with the L2 norm as in Equation (7).

\[
B_i = B_i / \sqrt{\|B_i\|^2 + \varepsilon}
\] (7)

| Value | V₁ | V₂ | V₃ | V₄ | V₅ | V₆ | V₇ | V₈ | V₉ |
|-------|----|----|----|----|----|----|----|----|----|
| Bins  | 0  | 20 | 40 | 60 | 80 | 100| 120| 140| 160|

**FIGURE 4** Representation of a 9-bins histogram
where, $\varepsilon$ is a small value added to the square of $B_i$ to avoid division by zero. The received code value is $1E - 05$. In this way, the necessary calculations are made for each defined grid boxes in the horizontal and vertical directions of the gradient matrix, and the HOG features are obtained.

4 PROPOSED METHODS

In this study, two ensemble learning models with different architectures are presented. The common features of both models are as follows: In order to increase classification accuracy of the dataset containing chest X-ray images, the capabilities of deep learning and machine learning algorithms are combined. The features of chest X-ray images were extracted by deep learning, and machine learning methods were used as classifiers. The majority voting (MV) technique was used so that the classifiers could decide together. The MV is an ensemble machine learning model that combines the predictions from multiple classifiers. In this technique, each classifier votes for a class, and the class with the most votes win. The ensemble’s anticipated target label is the mode of the distribution of individually predicted labels in statistical terms. It is a strategy that can be used to increase model performance, with the goal of outperforming any one model in the ensemble.

ResCNN, detailed in Figure 6, was used as a deep learning algorithm in the architecture of the proposed models. The ResCNN consists of six residual blocks and each block contains the Leaky ReLU activation function. In the classification block, there are max pooling 2D, fully connected and softmax layers. Training options of the ResCNN as follows: the solver for training network is set to Stochastic Gradient Descent with Momentum (SGDM) optimizer. Size of mini batches is 16 for each training iteration. Maximum number of epochs is 10 for the training process. Contribution of the gradient step, that is, momentum was set to 0.8. The number of epochs for dropping the learning rate was set to 5 and as the last parameter, factor for dropping the learning rate was set to 0.15.

The last layer of a neural network is a classification layer, which gathers the ultimate convoluted feature and produces a column vector with each row pointing to a class label. The probability estimation of each class label is represented by each member of the output vector. Traditional CNN methods end-up with a fully connected layer (FCL) in generally. In fact, the FCL is a holdover from the neural network concept of classification and feature extraction. In order to achieve the spatial transformation, the inputs are fully connected to the neurons to combine and reweight all the significant features (P01). Starting from this point of view, the features obtained from the FCL of the ResCNN, as shown in Figure 7, are used in training and testing of the proposed model #1 (PM1).

In the second phase, n machine learning methods are trained using the features extracted from ResCNN and predicted class labels are retrieved. Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), k-Nearest Neighbor (k-NN), Naïve Bayes (NB), and Decision Tree (DT) were used as machine learning algorithms in the architecture of the proposed models. The parameter settings of the machine learning algorithms used in experimental studies are shown in Table 2. These parameters were obtained by trial-and-error method depending on the extracted features with the LBP, HOG, and ResCNN.

In the HOG technique, the number of cells in a block was determined as 2 for all datasets used in the experiment. If a large block size value is chosen, it reduces the ability to suppress local lighting changes. Reducing the block size helps capture the importance of local pixels. While determining size of the HOG cell, it was gradually increased depending on the size of its images in the datasets. For $32 \times 32$, $64 \times 64$, $128 \times 128$, and $256 \times 256$ pixel datasets, cell sizes were determined as $[2 \times 2], [4 \times 4], [8 \times 8]$, and $[16 \times 16]$, respectively. Neighborhood Component Analysis (NCA) was used for the selection of features extracted using the HOG. Limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm was used.
for estimating feature weights in the NCA. The weights of the irrelevant features are close to zero. Therefore, the threshold value was set as $1E^{-6}$ for selecting features that are above the relative convergence tolerance on the gradient norm. The number of neighbors is kept constant at 16 to calculate the LBP for each pixel in the images for all datasets used in the experiment. In addition, the interpolation method used to calculate pixel neighbors is linear. The number of features extracted with LBP is $P(P - 1) + 3$, where $P$ represents the number of neighbors 16. Thus, 243 features were obtained from each image.
TABLE 2  Optimal parameter settings of machine learning algorithms

| Classifiers | Options                        |
|-------------|--------------------------------|
| LDA         | Discriminant type: Linear and pseudo-linear |
|             | Linear coefficient threshold: 0 |
| SVM         | Kernel function: Linear        |
|             | Kernel scale parameter: 1      |
|             | Prior probabilities: Empirical |
| NB          | Data distributions: Gaussian    |
| DT          | Maximum category levels: 10    |
|             | Min. number of leaf node: 3    |
|             | Min. number of branch node: 10 |
|             | Prior probabilities: Empirical |
| k-NN        | Number of neighbors: 7         |
|             | Distance metric: Euclidean     |
|             | Standardize: True              |

In the third and final phase, majority voting in Equation (9) is performed for the predictions of machine learning algorithms with combination in Equation (8). Let $S = \{\text{LDA, SVM, NB, DT, k-NN}\}$ be a set representing machine learning algorithms that can be used to solve a classification problem. Accordingly, $r$-element subsets (i.e., combinations without repetition) of the $S$:

$$\binom{S}{r} = \{M : |M| = r, M \subseteq S\}$$  \hspace{1cm} (8)

The number of machine learning algorithms used in this study is five ($|S| = 5$) and the value of $r$ is determined as 3. Thus, the number of 3-element subsets (M) of the $S$ is $C(|S|, r) = 10$. Where $C$ represents $r$-combinations of the $S$. The majority voting for the ensemble decision is:

$$f_{MV} = \arg\max \sum_{i=1}^{C(|S|, r)} \delta(\hat{y}, \text{mode}\{f_{p}(m_i), j = 1, 2, \ldots, r\}), \quad m_i \in M$$ \hspace{1cm} (9)

where, $\hat{y}$ is actual class label of a feature (or input) vector, $f_{p}(m_i)$ is the function that returns predicted class label of the $m_i$, and $\delta$ is the Kronecker Delta (KD) function in Equation (10).

$$\delta(x, y) =\begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases}, \quad \forall x, y \in \{\text{normal, covid, pneumonia}\}$$  \hspace{1cm} (10)

By performing the above processing steps, 10 ensemble learnings (M), each consisting of three different machine learning algorithms ($m_i$), are tested and used as classifier with the highest KD-score. Figure 8 shows the architecture of the proposed model #2 (PM2). The difference of this model compared to PM1 is that classifiers are supported by the LBP and HOG techniques.

In order to achieve higher classification accuracy, features were extracted from pre-processed chest X-ray images with the LBP and HOG techniques. Since the number of features obtained with HOG is quite large, NCA was applied to select some of them. The selection constraint was determined as the feature weights being greater than the relative convergence tolerance in the gradient norm. In the final stage, the classifiers were trained by combining the features obtained from LBP, HOG (selected by the NCA), and FCL of the ResCNN.

5  |  EXPERIMENTAL RESULTS

Experimental studies consist of four stages. In the first stage, the images in the chest X-ray dataset were sized in four different ways as $32 \times 32$, $64 \times 64$, $128 \times 128$, and $256 \times 256$ pixels. All images were then gray scale transformed and normalized, and pre-processing was completed.
After this process, the image dataset was partitioned with 10-fold cross-validation to measure the more realistic performance of the classifiers. The cross-validation is used to test model’s ability to predict new sample that was not used in predicting it, in order to flag problems like selection bias or overfitting. In addition, it is used to give an insight on how the model will generalize to a different dataset. In the classical approach, the accuracy of the model can be subjective as a result of manually partitioning the training and test data in certain proportions (e.g., 80% and 20%). For these reasons, it is aimed to obtain more objective and realistic results by partitioning dataset with the cross-validation technique.

Thus, 40 training/test data in different data sizes and combinations were applied to all classifiers. In the second stage, training and testing results were obtained for each cross-validated dataset using the ResCNN deep learning algorithm. In addition, 1000 features from ResCNN’s FCL were recorded for use in the proposed methods. In the third stage, five different machine learning algorithms were trained with the features obtained from ResCNN, called PM1 as shown in Figure 7 and tested using the majority voting technique with triple combinations. In the last stage, in addition to PM1, called PM2 as in Figure 8, the features obtained from the pre-processed images using the LBP and HOG techniques are combined with the features obtained from the ResCNN, and the training and testing of machine learning algorithms as in the PM1. The results were analyzed in detail.

All data in the dataset are labeled as Covid-19 (COV), Normal (NOR), and Pneumonia (PNE). All experiments were conducted using MATLAB version 2020b on a Core i7 CPU machine with 8 GB of RAM and 3.30 GHz. Table 3 shows the training and testing results of the ResCNN on the 10-fold cross-validation dataset. According to the results obtained, it is seen that the training and test scores gradually increase as the image sizes increase. Naturally, information loss occurs with the reduction of image sizes. In mean classification accuracies, the highest score was 96.77% for images in original image size, while the lowest score was 97.93% for 32 x 32 pixel images.
TABLE 3 Training and test accuracy of the ResCNN

| Folds | Training accuracy | Test accuracy |
|-------|-------------------|--------------|
|       | 32 × 32 | 64 × 64 | 128 × 128 | 256 × 256 | 32 × 32 | 64 × 64 | 128 × 128 | 256 × 256 |
| 1     | 92.41%  | 94.93%  | 96.79%  | 97.29%  | 93.49%  | 94.44%  | 95.79%  | 96.36%  |
| 2     | 93.02%  | 94.92%  | 96.48%  | 96.83%  | 92.73%  | 93.12%  | 96.18%  | 97.13%  |
| 3     | 93.32%  | 95.75%  | 96.28%  | 97.26%  | 92.93%  | 95.03%  | 95.60%  | 97.51%  |
| 4     | 92.86%  | 95.63%  | 96.34%  | 97.22%  | 93.12%  | 95.79%  | 96.18%  | 95.79%  |
| 5     | 93.32%  | 94.87%  | 96.77%  | 97.31%  | 90.44%  | 91.97%  | 94.65%  | 95.98%  |
| 6     | 93.06%  | 95.51%  | 96.14%  | 97.23%  | 90.63%  | 94.65%  | 93.88%  | 95.98%  |
| 7     | 93.05%  | 95.60%  | 95.73%  | 96.99%  | 89.10%  | 93.31%  | 95.98%  | 96.75%  |
| 8     | 91.92%  | 95.87%  | 96.30%  | 97.35%  | 91.59%  | 95.22%  | 96.56%  | 96.56%  |
| 9     | 93.48%  | 94.99%  | 96.12%  | 97.23%  | 91.78%  | 93.69%  | 96.37%  | 97.51%  |
| 10    | 92.68%  | 95.23%  | 96.42%  | 97.03%  | 93.49%  | 95.21%  | 96.36%  | 98.08%  |
| Mean  | 92.91%  | 95.33%  | 96.34%  | 97.18%  | 91.93%  | 94.24%  | 95.75%  | 96.77%  |

FIGURE 9 Mean classification accuracy of the PM1, ResCNN and machine learning methods

With the ResCNN, average training times of the dataset are 382, 785, 2414, and 9975 s for 32 × 32, 64 × 64, 128 × 128, and 256 × 256 image sizes, respectively. Figure 9 shows the mean classification accuracies of the PM1, ResCNN and machine learning algorithms according to different image sizes. In Table 4, the classification results of the same methods with 10-fold cross-validation are given in detail.

As a result of the experimental studies, it was seen that PM1 slightly improved the classification accuracy compared to ResCNN for all image sizes. The machine learning combination that won the most in the majority voting was the trio of the k-NN, LDA, and SVM.

From Tables 5–8, classification accuracies of machine learning algorithms, ResCNN and PM2 are given on a scaled and 10-fold cross-validation dataset of 32 × 32, 64 × 64, 128 × 128, and 256 × 256 image sizes, respectively. Inspecting the results of these tables shows that PM2 has the best classification accuracy in almost all cross-validation data. With the inclusion of features obtained from LBP and HOG techniques in the PM2 model, it is clearly seen that a higher score is obtained than PM1. In addition, LDA, SVM, and k-NN classifiers achieved better results than ResCNN.

Figure 10 shows the classification results of ResCNN, PM1, and PM2 according to the different image sizes of the dataset. The values between 1 and 10 on the horizontal axes in the charts indicate cross-validation, and on the vertical axes, classification accuracies are shown as percentages, with different colors and shapes for each method. The performance of PM2 on all cross-validation data on 32 × 32 pixel images in Figure 10A is considerably higher than the others. While the average classification accuracy of PM2 on 32 × 32 pixel images is 95.66%, ResCNN can only exceed this rate with 128 × 128 pixel images.

To measure the performance of the PM2 according to the class labels, four measurements, namely accuracy, sensitivity, specificity and precision, were recorded. According to the results obtained, it was observed that the COV label was stable with an accuracy rate of 98.7% in all measurements. When we look at the precision rates, it is 99% in 256 × 256 pixel images and 98% in the others. Precision value is especially important when the cost of False Positive (FP) estimation is high. For example, if the patient who should be labeled as NOR in the radiological diagnosis is labeled as COV or PNE (FP), then it will weaken the reliability of the system by saying that it is a sign of infection for healthy people. For this reason, a high precision
| Fold      | Machine learning methods | ResCNN | PM1 | Combinations |
|-----------|--------------------------|--------|-----|--------------|
|           | LDA          | SVM    | NB   | DT           | k-NN |            |
| 32 × 32   | 1            | 94.06% | 92.72% | 79.50% | 79.12% | 92.34% |
|           | 2            | 93.31% | 93.50% | 86.42% | 80.12% | 89.29% |
|           | 3            | 92.93% | 92.54% | 74.19% | 78.59% | 93.69% |
|           | 4            | 93.12% | 93.31% | 78.20% | 80.69% | 88.91% |
|           | 5            | 90.44% | 92.35% | 78.20% | 74.19% | 87.95% |
|           | 6            | 90.63% | 91.97% | 80.03% | 73.04% | 90.06% |
|           | 7            | 91.40% | 89.29% | 77.06% | 75.91% | 88.34% |
|           | 8            | 91.21% | 91.16% | 76.86% | 80.31% | 90.06% |
|           | 9            | 89.68% | 91.78% | 81.45% | 77.44% | 91.21% |
|           | 10           | 93.10% | 92.15% | 85.82% | 76.63% | 88.70% |
| 64 × 64   | 1            | 94.06% | 93.87% | 87.36% | 77.78% | 86.97% |
|           | 2            | 93.69% | 94.26% | 80.88% | 79.16% | 90.06% |
|           | 3            | 94.26% | 94.26% | 84.13% | 77.25% | 89.29% |
|           | 4            | 93.31% | 95.41% | 89.10% | 80.50% | 91.40% |
|           | 5            | 91.40% | 93.88% | 76.48% | 78.01% | 87.57% |
|           | 6            | 91.59% | 92.16% | 83.56% | 82.41% | 90.06% |
|           | 7            | 91.12% | 93.88% | 85.66% | 80.50% | 90.25% |
|           | 8            | 91.40% | 95.22% | 79.73% | 80.50% | 91.21% |
|           | 9            | 93.49% | 94.83% | 88.51% | 82.57% | 92.15% |
|           | 10           | 93.10% | 92.15% | 85.82% | 76.63% | 88.70% |
| 128 × 128 | 1            | 94.44% | 96.36% | 89.66% | 86.78% | 92.91% |
|           | 2            | 95.22% | 95.41% | 92.93% | 82.22% | 96.75% |
|           | 3            | 95.22% | 97.32% | 90.63% | 83.94% | 96.75% |
|           | 4            | 96.37% | 95.79% | 91.78% | 84.51% | 95.60% |
|           | 5            | 94.84% | 95.41% | 88.72% | 80.88% | 93.31% |
|           | 6            | 92.73% | 94.26% | 87.38% | 85.47% | 93.12% |
|           | 7            | 95.41% | 96.75% | 90.06% | 79.92% | 93.31% |
|           | 8            | 95.41% | 97.13% | 95.60% | 89.09% | 95.41% |
|           | 9            | 94.46% | 96.37% | 90.63% | 80.69% | 95.41% |
|           | 10           | 96.36% | 96.94% | 93.87% | 85.06% | 94.06% |
| 256 × 256 | 1            | 96.55% | 97.32% | 90.04% | 90.04% | 95.59% |
|           | 2            | 95.60% | 97.13% | 95.60% | 90.82% | 94.84% |
|           | 3            | 96.18% | 97.71% | 96.56% | 92.93% | 95.99% |
|           | 4            | 95.79% | 96.37% | 95.60% | 89.48% | 95.41% |
|           | 5            | 96.18% | 96.94% | 95.41% | 89.87% | 96.37% |
|           | 6            | 94.84% | 96.94% | 95.60% | 90.06% | 95.22% |
|           | 7            | 95.79% | 96.37% | 95.60% | 89.87% | 96.37% |
|           | 8            | 95.79% | 96.94% | 96.75% | 90.63% | 95.41% |
|           | 9            | 97.13% | 97.71% | 96.94% | 87.00% | 96.75% |
|           | 10           | 95.59% | 96.94% | 87.74% | 88.89% | 96.74% |
### TABLE 5  Classification accuracy of the PM2, ResCNN, and machine learning methods by $32 \times 32$ image size (highest scores highlighted in bold)

| Fold | LDA (%) | SVM (%) | NB (%) | DT (%) | k-NN (%) | ResCNN (%) | PM2 (%) | Combinations |
|------|---------|---------|--------|--------|----------|------------|---------|--------------|
| 32 x 32 |         |         |        |        |          |            |         |              |
| 1    | 95.98   | 93.68   | 88.12  | 82.57  | 95.40    | 93.49      | 96.36   | (kNN, LDA, SVM) |
| 2    | 95.22   | 93.50   | 90.44  | 79.92  | 94.26    | 92.73      | 96.18   | (DT, kNN, LDA) |
| 3    | 96.56   | 94.07   | 83.37  | 78.59  | 94.07    | 92.93      | 96.56   | (kNN, LDA, SVM) |
| 4    | 94.84   | 93.88   | 86.62  | 85.28  | 94.66    | 93.12      | 96.56   | (DT, kNN, LDA) |
| 5    | 93.88   | 92.54   | 87.19  | 81.84  | 91.97    | 90.44      | 94.46   | (kNN, LDA, SVM) |
| 6    | 94.46   | 92.35   | 87.00  | 77.63  | 93.88    | 90.63      | 95.41   | (kNN, LDA, SVM) |
| 7    | 93.88   | 90.25   | 83.75  | 80.69  | 92.73    | 89.10      | 94.26   | (kNN, LDA, SVM) |
| 8    | 95.60   | 93.12   | 89.29  | 84.70  | 94.46    | 91.59      | 95.99   | (kNN, LDA, SVM) |
| 9    | 93.69   | 91.59   | 87.38  | 80.69  | 95.60    | 91.78      | 95.03   | (kNN, LDA, NB) |
| 10   | 95.21   | 92.72   | 91.19  | 81.99  | 93.68    | 93.49      | 95.79   | (kNN, LDA, SVM) |

Mean CACC: 94.93% 92.77% 87.43% 81.39% 94.05% 91.93% 95.66%

### TABLE 6  Classification accuracy of the PM2, ResCNN, and machine learning methods by $64 \times 64$ image size (highest scores highlighted in bold)

| Fold | LDA (%) | SVM (%) | NB (%) | DT (%) | k-NN (%) | ResCNN (%) | PM2 (%) | Combinations |
|------|---------|---------|--------|--------|----------|------------|---------|--------------|
| 64 x 64 |         |         |        |        |          |            |         |              |
| 1    | 95.21   | 93.87   | 91.00  | 79.31  | 93.30    | 94.44      | 95.79   | (kNN, LDA, NB) |
| 2    | 95.99   | 94.07   | 88.34  | 82.60  | 95.79    | 93.12      | 96.56   | (kNN, LDA, SVM) |
| 3    | 98.09   | 94.84   | 86.81  | 84.32  | 92.54    | 95.03      | 97.13   | (kNN, LDA, SVM) |
| 4    | 94.46   | 95.79   | 91.78  | 83.94  | 95.99    | 95.79      | 96.56   | (kNN, NB, SVM) |
| 5    | 91.97   | 93.69   | 85.85  | 81.64  | 93.31    | 91.97      | 95.60   | (kNN, LDA, SVM) |
| 6    | 93.88   | 92.35   | 87.76  | 80.88  | 93.12    | 94.65      | 94.84   | (DT, kNN, LDA) |
| 7    | 94.65   | 93.88   | 90.25  | 81.84  | 93.88    | 93.31      | 95.41   | (kNN, LDA, NB) |
| 8    | 95.41   | 95.79   | 91.01  | 82.22  | 96.37    | 95.22      | 96.94   | (kNN, NB, SVM) |
| 9    | 94.65   | 95.41   | 83.94  | 84.90  | 92.54    | 93.69      | 96.18   | (kNN, LDA, SVM) |
| 10   | 94.83   | 95.02   | 91.19  | 84.10  | 95.21    | 95.21      | 95.79   | (kNN, LDA, SVM) |

Mean CACC: 94.91% 94.47% 88.79% 82.57% 94.20% 94.24% 96.08%

### TABLE 7  Classification accuracy of the PM2, ResCNN, and machine learning methods by $128 \times 128$ image size (highest scores highlighted in bold)

| Fold | LDA (%) | SVM (%) | NB (%) | DT (%) | k-NN (%) | ResCNN (%) | PM2 (%) | Combinations |
|------|---------|---------|--------|--------|----------|------------|---------|--------------|
| 128 x 128 |         |         |        |        |          |            |         |              |
| 1    | 69.73   | 96.36   | 92.91  | 87.17  | 95.02    | 95.79      | 96.55   | (kNN, NB, SVM) |
| 2    | 96.94   | 95.41   | 93.50  | 84.70  | 96.75    | 96.18      | 97.32   | (kNN, LDA, SVM) |
| 3    | 96.56   | 97.51   | 90.63  | 88.15  | 96.94    | 95.60      | 97.90   | (kNN, LDA, SVM) |
| 4    | 97.32   | 95.79   | 91.59  | 88.72  | 96.37    | 96.18      | 97.51   | (DT, kNN, LDA) |
| 5    | 85.66   | 95.41   | 89.48  | 84.32  | 94.46    | 94.65      | 95.22   | (kNN, NB, SVM) |
| 6    | 87.95   | 94.46   | 90.82  | 85.09  | 95.79    | 93.88      | 96.56   | (kNN, NB, SVM) |
| 7    | 92.93   | 96.75   | 93.50  | 87.00  | 96.56    | 95.98      | 97.51   | (kNN, LDA, NB) |
| 8    | 95.79   | 97.13   | 92.54  | 86.42  | 97.32    | 96.56      | 97.32   | (kNN, NB, SVM) |
| 9    | 90.44   | 96.37   | 92.16  | 83.37  | 96.18    | 96.37      | 97.51   | (kNN, LDA, SVM) |
| 10   | 95.98   | 96.74   | 92.72  | 90.81  | 96.17    | 96.36      | 97.89   | (kNN, LDA, SVM) |

Mean CACC: 90.93% 96.19% 91.99% 86.57% 96.16% 95.75% 97.13%
TABLE 8 Classification accuracy of the PM2, ResCNN, and machine learning methods by 256 x 256 image size (highest scores highlighted in bold)

| Fold | Machine learning methods | LDA | SVM | NB | DT | k-NN | ResCNN | PM2 | Combinations |
|------|--------------------------|-----|-----|----|----|------|--------|-----|--------------|
| 256 x 256 | 1 | 92.34% | 97.13% | 95.79% | 90.04% | 97.13% | 96.36% | 98.28% | (kNN, LDA, SVM) |
| | 2 | 96.75% | 97.13% | 95.99% | 91.40% | 97.71% | 97.13% | 98.47% | (kNN, LDA, SVM) |
| | 3 | 96.37% | 97.71% | 95.79% | 93.12% | 97.71% | 97.51% | 98.47% | (kNN, NB, SVM) |
| | 4 | 93.69% | 96.37% | 94.65% | 88.53% | 95.79% | 95.79% | 97.51% | (kNN, LDA, SVM) |
| | 5 | 97.32% | 96.94% | 96.37% | 91.40% | 97.90% | 95.98% | 98.09% | (DT, kNN, LDA) |
| | 6 | 96.18% | 96.94% | 95.99% | 90.44% | 97.13% | 95.98% | 97.51% | (DT, kNN, LDA) |
| | 7 | 92.73% | 96.37% | 94.84% | 90.44% | 96.75% | 96.75% | 97.13% | (kNN, LDA, SVM) |
| | 8 | 96.56% | 96.94% | 96.37% | 90.63% | 98.28% | 96.56% | 97.51% | (kNN, LDA, NB) |
| | 9 | 96.94% | 97.71% | 95.99% | 87.76% | 97.71% | 97.51% | 98.09% | (kNN, NB, SVM) |
| | 10 | 96.17% | 96.94% | 92.91% | 92.53% | 96.55% | 98.08% | 98.08% | (kNN, LDA, SVM) |

Mean CACC: 95.50% 97.02% 95.47% 90.63% 97.26% 96.77% 97.92%

FIGURE 10 Comparison of classification accuracies of the PM1, PM2, and ResCNN with 10-fold cross validation

value is an important criterion in model selection. From this point of view, it is seen that the PM2 can achieve the desired precision value in smaller images. Another interesting point is that specificity score of the COV label in Table 9 is over 99%. Specificity indicates how many of the patients with True Negative (TN) status were correctly detected. However, as the images sizes get larger (i.e., from Tables 10-12), it is seen that this success in the specificity rate passes to the NOR label.
TABLE 9  Test results of the PM2 by dataset with 32 × 32 image size

| Fold | Accuracy | Sensitivity | Specificity | Precision |
|------|----------|-------------|-------------|-----------|
|      | COV      | NOR         | PNE         | COV       | NOR       | PNE         | COV       | NOR       | PNE         | COV       | NOR       | PNE         |
| 32 × 32 | 1 99.0% | 96.9% | 96.7% | 97.5% | 95.0% | 96.7% | 99.7% | 98.0% | 96.8% | 99.4% | 96.1% | 94.1% |
|        | 2 99.2% | 96.6% | 96.6% | 98.8% | 96.1% | 93.9% | 99.4% | 96.8% | 98.0% | 98.8% | 94.0% | 96.0% |
|        | 3 98.7% | 97.3% | 97.1% | 98.8% | 97.8% | 93.3% | 98.6% | 97.1% | 99.1% | 97.0% | 94.6% | 98.2% |
|        | 4 98.9% | 97.5% | 96.8% | 98.2% | 96.7% | 95.0% | 99.2% | 98.0% | 97.7% | 98.2% | 96.1% | 95.5% |
|        | 5 98.3% | 95.2% | 95.4% | 96.9% | 95.0% | 91.7% | 98.9% | 95.3% | 97.4% | 97.5% | 91.4% | 94.8% |
|        | 6 98.7% | 95.8% | 96.4% | 98.2% | 94.4% | 93.9% | 98.9% | 96.5% | 97.7% | 97.6% | 93.4% | 95.5% |
|        | 7 98.9% | 95.0% | 94.6% | 97.5% | 94.4% | 91.1% | 99.4% | 95.3% | 96.5% | 98.8% | 91.4% | 93.2% |
|        | 8 98.7% | 96.8% | 96.6% | 98.1% | 96.1% | 93.9% | 98.9% | 97.7% | 98.8% | 98.2% | 94.1% | 97.7% |
|        | 9 98.1% | 95.8% | 96.2% | 96.9% | 96.7% | 91.7% | 98.6% | 95.3% | 98.5% | 96.9% | 91.6% | 97.1% |
|        | 10 98.9% | 96.6% | 96.2% | 98.1% | 92.8% | 96.7% | 99.2% | 98.5% | 95.9% | 98.1% | 97.1% | 92.6% |
| Mean   | 98.7% | 96.3% | 96.3% | 97.9% | 95.5% | 93.8% | 99.1% | 96.8% | 97.5% | 98.0% | 94.0% | 95.3% |

TABLE 10  Test results of the PM2 by dataset with 64 × 64 image size

| Fold | Accuracy | Sensitivity | Specificity | Precision |
|------|----------|-------------|-------------|-----------|
|      | COV      | NOR         | PNE         | COV       | NOR       | PNE         | COV       | NOR       | PNE         | COV       | NOR       | PNE         |
| 64 × 64 | 1 98.7% | 96.2% | 96.7% | 98.8% | 92.2% | 96.7% | 98.6% | 98.2% | 96.8% | 97.0% | 96.5% | 94.1% |
|        | 2 99.0% | 97.1% | 96.9% | 98.8% | 97.8% | 93.3% | 99.2% | 96.8% | 98.8% | 98.2% | 94.1% | 97.7% |
|        | 3 99.4% | 97.7% | 97.1% | 99.4% | 98.9% | 93.3% | 99.4% | 97.1% | 99.1% | 98.8% | 94.7% | 98.2% |
|        | 4 99.2% | 96.9% | 96.9% | 99.4% | 95.6% | 95.0% | 99.2% | 97.7% | 98.0% | 98.2% | 95.6% | 96.1% |
|        | 5 99.4% | 96.0% | 95.8% | 99.4% | 97.8% | 90.0% | 99.4% | 95.0% | 98.8% | 98.8% | 91.2% | 97.6% |
|        | 6 97.9% | 95.8% | 96.0% | 96.3% | 92.2% | 96.1% | 98.6% | 97.7% | 95.9% | 96.9% | 95.4% | 92.5% |
|        | 7 98.9% | 96.2% | 95.8% | 96.9% | 95.0% | 94.4% | 99.7% | 96.8% | 96.5% | 99.4% | 94.0% | 93.4% |
|        | 8 98.7% | 97.7% | 97.5% | 99.4% | 96.1% | 95.6% | 98.3% | 98.5% | 98.5% | 96.4% | 97.2% | 97.2% |
|        | 9 98.5% | 96.9% | 96.9% | 97.5% | 96.7% | 94.4% | 98.9% | 97.1% | 98.3% | 97.5% | 94.6% | 96.6% |
|        | 10 99.2% | 96.2% | 96.2% | 98.1% | 93.9% | 95.6% | 99.7% | 97.4% | 96.5% | 99.4% | 94.9% | 93.5% |
| Mean   | 98.7% | 98.9% | 96.7% | 96.6% | 98.4% | 95.6% | 94.4% | 99.1% | 97.2% | 97.7% | 98.0% | 94.8% |

In addition to the model classification performance criteria, multiclass ROC curves were prepared for both PM1 and PM2 according to different image sizes (32 × 32, 64 × 64, 128 × 128, and 256 × 256) as shown in Figure 11. In addition, AUC (Area Under the ROC Curve) scores were calculated separately for each class in subfigures. The AUC_C, AUC_N, and AUC_P labels on this figure represent AUC-scores of Covid19, Normal, and Pneumonia, respectively. The AUC-scores for each model increase in proportion to the image sizes. Therefore, the lowest scores are on 32 × 32 pixel images, and the highest scores are on 256 × 256 pixel images. The most striking point in these figures is that AUC-scores of the ResCNN with 128 × 128 pixel images, the PM2 was able to obtain with 32 × 32 pixel images.

6  DISCUSSION AND CONCLUSION

The CNNs can extract features of objects by scanning images with convolution kernels at a significantly reduced computational cost. It also increases the resistance of network to spatial variations thanks to the pooling layer. The features obtained by the CNN tend to be abstract that are difficult to
### TABLE 11 Test results of the PM2 by dataset with 128 × 128 image size

| Fold | Accuracy COV | Accuracy NOR | Accuracy PNE | Sensitivity COV | Sensitivity NOR | Sensitivity PNE | Specificity COV | Specificity NOR | Specificity PNE | Precision COV | Precision NOR | Precision PNE |
|------|--------------|--------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------|--------------|--------------|
| 128 × 128 | 98.9% | 97.3% | 96.9% | 97.5% | 96.7% | 95.6% | 99.4% | 97.7% | 97.7% | 98.8% | 95.6% | 95.6% |
| 2   | 99.2% | 97.3% | 98.1% | 99.4% | 96.1% | 96.7% | 99.2% | 98.0% | 98.8% | 98.2% | 96.1% | 97.8% |
| 3   | 99.6% | 98.3% | 97.9% | 98.8% | 97.2% | 97.8% | 100.0% | 98.8% | 98.0% | 100.0% | 97.8% | 96.2% |
| 4   | 99.8% | 97.7% | 97.5% | 99.4% | 97.2% | 96.1% | 100.0% | 98.0% | 98.3% | 100.0% | 96.2% | 96.6% |
| 5   | 99.0% | 95.8% | 95.6% | 98.2% | 96.7% | 91.1% | 99.4% | 95.3% | 98.0% | 98.8% | 91.6% | 95.9% |
| 6   | 98.9% | 97.3% | 96.9% | 97.5% | 97.2% | 95.0% | 99.4% | 97.4% | 98.0% | 98.8% | 95.1% | 97.8% |
| 7   | 99.0% | 98.1% | 97.9% | 98.8% | 98.3% | 95.6% | 99.2% | 98.0% | 99.1% | 98.8% | 96.2% | 98.3% |
| 8   | 99.2% | 98.1% | 97.3% | 99.4% | 97.8% | 95.0% | 99.2% | 98.2% | 98.5% | 98.2% | 96.7% | 97.2% |
| 9   | 98.7% | 98.5% | 97.9% | 98.1% | 98.9% | 95.6% | 98.9% | 98.2% | 99.1% | 97.5% | 96.8% | 98.3% |
| 10  | 99.4% | 98.3% | 98.1% | 98.1% | 97.8% | 97.8% | 100.0% | 98.5% | 98.2% | 100.0% | 97.2% | 96.7% |

**Mean**

| Accuracy COV | Accuracy NOR | Accuracy PNE | Sensitivity COV | Sensitivity NOR | Sensitivity PNE | Specificity COV | Specificity NOR | Specificity PNE | Precision COV | Precision NOR | Precision PNE |
|--------------|--------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------|--------------|--------------|
| 98.7% | 99.2% | 97.7% | 97.4% | 98.5% | 97.4% | 95.6% | 99.5% | 97.8% | 98.4% | 98.8% | 95.9% |

### TABLE 12 Test results of the PM2 by dataset with 256 × 256 image size

| Fold | Accuracy COV | Accuracy NOR | Accuracy PNE | Sensitivity COV | Sensitivity NOR | Sensitivity PNE | Specificity COV | Specificity NOR | Specificity PNE | Precision COV | Precision NOR | Precision PNE |
|------|--------------|--------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------|--------------|--------------|
| 256 × 256 | 99.4% | 98.5% | 98.7% | 98.8% | 97.8% | 98.3% | 99.7% | 98.8% | 98.8% | 99.4% | 97.8% | 97.8% |
| 2   | 99.8% | 98.5% | 98.7% | 100.0% | 98.3% | 97.2% | 99.7% | 98.5% | 99.4% | 99.4% | 97.3% | 98.9% |
| 3   | 99.6% | 98.9% | 98.5% | 100.0% | 98.9% | 96.7% | 99.4% | 98.8% | 99.4% | 98.8% | 97.8% | 98.9% |
| 4   | 99.2% | 98.1% | 97.7% | 98.2% | 98.9% | 95.6% | 99.7% | 97.7% | 98.8% | 99.4% | 95.7% | 97.7% |
| 5   | 99.2% | 98.3% | 98.7% | 99.4% | 97.8% | 97.2% | 99.2% | 98.5% | 99.4% | 98.2% | 97.2% | 98.9% |
| 6   | 99.0% | 97.9% | 98.1% | 99.4% | 97.2% | 96.1% | 98.9% | 98.3% | 99.1% | 97.6% | 96.7% | 98.3% |
| 7   | 99.6% | 97.3% | 97.3% | 99.4% | 97.2% | 95.0% | 99.7% | 97.4% | 98.5% | 99.4% | 95.1% | 97.2% |
| 8   | 98.9% | 98.5% | 97.7% | 99.4% | 98.3% | 95.0% | 98.6% | 98.5% | 99.1% | 97.0% | 97.3% | 98.3% |
| 9   | 99.6% | 98.5% | 98.1% | 100.0% | 100.0% | 94.4% | 99.4% | 97.7% | 100.0% | 98.8% | 95.8% | 100.0% |
| 10  | 99.4% | 98.7% | 98.1% | 98.8% | 98.3% | 97.2% | 99.7% | 98.8% | 98.5% | 99.4% | 97.8% | 97.2% |

**Mean**

| Accuracy COV | Accuracy NOR | Accuracy PNE | Sensitivity COV | Sensitivity NOR | Sensitivity PNE | Specificity COV | Specificity NOR | Specificity PNE | Precision COV | Precision NOR | Precision PNE |
|--------------|--------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------|--------------|--------------|
| 98.7% | 99.4% | 98.3% | 98.1% | 99.3% | 98.3% | 96.3% | 99.4% | 98.3% | 99.1% | 98.7% | 96.8% |

Machine learning algorithms need features that it can use to perform a task. Feature extraction is usually quite complex and requires detailed knowledge of the problem domain. In this respect, it is disadvantageous compared to CNNs. In this study, the disadvantageous situation of machine
FIGURE 11  ROC curves and AUC scores of the ResCNN, PM1, and PM2
TABLE 13  Comparison of accuracy scores of the proposed models and ResCNN on the diagnosis of Covid-19

| Datasets                      | Methods  | Test accuracies          |
|-------------------------------|----------|--------------------------|
|                               |          | Overall | Highest | Lowest |
| Chest X-ray (32 × 32 pixels)  | ResCNN   | 91.93% | 93.49% | 89.10% |
|                               | PM1      | 93.15% | 94.46% | 91.59% |
|                               | PM2      | 95.66% | 96.56% | 94.26% |
| Chest X-ray (64 × 64 pixels)  | ResCNN   | 94.24% | 95.79% | 91.97% |
|                               | PM1      | 94.61% | 96.75% | 92.54% |
|                               | PM2      | 96.08% | 97.13% | 94.84% |
| Chest X-ray (128 × 128 pixels)| ResCNN   | 95.76% | 96.56% | 93.88% |
|                               | PM1      | 96.48% | 97.51% | 94.46% |
|                               | PM2      | 97.13% | 97.90% | 95.22% |
| Chest X-ray (256 × 256 pixels)| ResCNN   | 96.77% | 98.08% | 95.79% |
|                               | PM1      | 97.15% | 97.90% | 96.37% |
|                               | PM2      | 97.92% | 98.47% | 97.13% |

TABLE 14  Total execution time of the methods (in seconds)

| Datasets                      | MLAs  | HOG + LBP | ResCNN | PM1   | PM2   |
|-------------------------------|-------|-----------|--------|-------|-------|
| Chest X-ray (32 × 32 pixels)  | 125   | 900       | 3819   | 3944  | 4844  |
| Chest X-ray (64 × 64 pixels)  | 147   | 964       | 7809   | 7956  | 8920  |
| Chest X-ray (128 × 128 pixels)| 166   | 1057      | 24,143 | 24,309| 25,366|
| Chest X-ray (256 × 256 pixels)| 198   | 1633      | 99,745 | 99,943| 101,576|

learning algorithms is eliminated by using the features derived from the ResCNN. Thus, by combining the predictions of many machine learning algorithms with majority voting techniques, it has been transformed into an ensemble learning model, which can achieve higher classification accuracy.

If we consider the performance of the proposed models in terms of both time cost and classification accuracy; the classification accuracy on the ResCNN’s 128 × 128 pixel dataset shown in Table 13 is 95.76% and the time spent on training the model is 24,143 s (6h 42min), as can be seen in Table 14. The PM2, with its 32 × 32 pixel dataset, achieved a score very close to the classification accuracy of the ResCNN’s 128 × 128 pixel dataset. Looking at the total time spent for this in Table 14, it is seen that it is 4844 s (1h and 20min). This shows that PM2 can achieve the same classification accuracy in (about five times) shorter than the ResCNN. The total time spent for 96.77% classification accuracy achieved by the ResCNN on a 256 × 256 pixel dataset is 99,745 s (27h 42min). The PM2, on the other hand, has a classification accuracy of 96.08% with a 64 × 64 pixel dataset and 97.13% with a 128 × 128 pixel dataset. We can interpret this as follows; For the PM2 to achieve 96.77% classification accuracy (as the closest value), the image dimensions in the dataset are expected to have a value greater than 64 and less than 128. That is, the PM2’s runtime is between 8920 s (2h 28min) and 25,366 s (7h 2min). However, since we cannot calculate this precisely with the information we have, the PM2 shows that in the worst case, it can achieve higher classification accuracy than the ResCNN in (about four times) shorter time.

Although the classification performance of the PM1 was slightly better than the ResCNN, it lagged behind the PM2. In the PM1 model, machine learning algorithms are trained using features obtained entirely from ResCNN. Although the majority voting technique slightly improved the results, it did not achieve a higher classification score as in the PM2. The reason why the PM2 gives better results than the PM1 is undoubtedly the inclusion of the HOG and LBP extracted features. Thus, by further enriching the features produced by the ResCNN, it strengthened the distinctive ability of the model. The PM2 was applied to the chest X-ray image dataset and the obtained classification results proved to be significantly successful. In addition, it has been proven by experimental studies that reduces the running time cost between four and five times in order to achieve classification accuracy of the ResCNN. It is hoped that the proposed method will improve Covid-19 detection and lead to more reliable radiology systems.
CONFLICT OF INTEREST
The author declares that there is no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in Chest X-Ray Images Database at https://www.kaggle.com/sachinkumar413/covid-pneumonia-normal-chest-xray-images, reference 28.

ORCID
Abdullah Elen https://orcid.org/0000-0003-1644-0476

REFERENCES
1. Dwivedi DK, Parihar A, Kohli N, Dandu HR, Saxena SK. Radiological perspective of the novel coronavirus disease 2019 (COVID-19). In: Chandra P, Roy S, eds. Diagnostic Strategies for COVID-19 and Other Coronaviruses. Medical Virology: From Pathogenesis to Disease Control. Springer; 2020:37-49. doi:10.1007/978-81-5-6006-4-3
2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-733. doi:10.1056/nejmoa2001217
3. Han T, Ryu B, Lee S, et al. Management following the first confirmed case of SARS-CoV-2 in a domestic cat associated with a massive outbreak in South Korea. One Health. 2021;13:100328. doi:10.1016/j.onehlt.2021.100328
4. Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. Int J Infect Dis. 2020;94:44-48. doi:10.1016/j.ijid.2020.03.004
5. Cyranoski D. Profile of a killer: the complex biology powering the coronavirus pandemic. Nature. 2020;581(7806):22-27. doi:10.1038/d41586-020-01315-7
6. Lai C-C, Liu YH, Wang CY, et al. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): facts and myths. J Microbiol Immunol Infect. 2020;53(3):404-412. doi:10.1016/j.jmii.2020.02.012
7. Azer SA, COVID-19: pathophysiology, diagnosis, complications and investigational therapeutics. New Microb New Infect. 2020;37:100738. doi:10.1016/j.jnmni.2020.100738
8. Jayaweera M, Perera H, Gunawardana B, Manutunge J. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. Environ Res. 2020;188:109819. doi:10.1016/j.envres.2020.109819
9. Naveed M, Naeeem M, ur Rahman M, et al. Review of potential risk groups for coronavirus disease 2019 (COVID-19). New Microb New Infect. 2021;41:100849. doi:10.1016/j.jnmni.2021.100849
10. Turkish Ministry of Health. COVID-19 Information Platform; 2021. https://covid19.saglik.gov.tr/TR-66300/covid-19-nedir-.html
11. Saygılı A. A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods. Appl Soft Comput. 2021;105:107323. doi:10.1016/j.asoc.2021.107323
12. Bahadur Chandra T, Verma K, Kumar Singh B, Jain D, Singh Netam S. Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble. Expert Syst Appl. 2020;165:114054. doi:10.1016/j.eswa.2020.114054
13. Hira S, Bai A, Hira S. An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images. Appl Intell. 2021;51:2864-2889. doi:10.1007/s10489-020-00865-4
14. Ozturk T, Talo M, Yıldırım EA, Baloglu UB, Yıldırım Ö, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020;121:103792. doi:10.1016/j.compbiomed.2020.103
15. Chowdhury NK, Rahman MM, Kabir MA. PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images. Health Inform Sci Syst. 2020;8:27. doi:10.1007/j.ijmedinf.2020.06042
16. Heidari M, Mirmaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. Int J Med Inform. 2020;144:104284. doi:10.1016/j.ijmedinf.2020.10428
17. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med. 2020;43:635-640. doi:10.1016/j.ijeswa.2020.12466-020-00865-4
18. Chouhan V, Singh SK, Khamparia A, et al. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci. 2020;10(2):559. doi:10.3390/app10020559
19. Motamed S, Rogalla P, Khalvati F. RANDGAN: randomized generative adversarial network for detection of COVID-19 in chest X-ray. Sci Rep. 2021;11:8602. doi:10.1038/s41598-021-87994-2
20. Goel T, Murugan R, Mirjalili S, Chakrabarty DK. Automatic screening of COVID-19 using an optimized generative adversarial network. Cogn Comput. 2021;1-16. doi:10.1007/12255-20-00978-5
21. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296(2):E32-E40. doi:10.1148/radiol.2020200642
22. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjuman F, Pinheiro PR. CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. IEEE Access. 2020;8:91916-91923. doi:10.1109/access.2020.2994762
23. Ismael AM, Şengür A. Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Syst Appl. 2021;164:114054. doi:10.1016/j.eswa.2021.114054
24. Hussain E, Hasan M, Rahman MA, Lee I, Tamanna T, Parvez MZ. CoroDet: a deep learning based classification for COVID-19 detection using chest X-ray images. Chaos Solit Fract. 2021;142:110495. doi:10.1016/j.chaos.2020.110495
25. Chandra TB, Verma K, Kumar Singh B, Jain D, Singh Netam S. Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble. Expert Syst Appl. 2020;165:113909. doi:10.1016/j.eswa.2020.113909
26. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (COVID-19) classification using CT images by machine learning methods. arXiv:2003.09424; 2020.
27. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Appl. 2021;24:1207-1220. doi:10.1007/s10044-021-00984-y
28. Kaggle. Chest X-Ray Images Database; 2021. https://www.kaggle.com/sachinkumar413/covid-pneumonia-normal-chest-xray-images
29. Liu P-Y, Li Z-M. A feature extraction method based on local binary pattern preprocessing and wavelet transform. *Int J Pattern Recog Artif Intellig*. 2020;34(13):2050030. doi:10.1142/s0218001420500305
30. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn*. 1996;29(2):51-59. doi:10.1016/0031-3203(95)00067-4
31. Bhagat PK, Choudhary P, Singh KM. A comparative study for brain tumor detection in MRI images using texture features. *Sens Health Monit*. 2019;5:259-287. doi:10.1016/b978-0-12-819361-7.00013-0
32. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), Vol. 1; 2005:886-893. doi:10.1109/CVPR.2005.177
33. Lee WY, Ko KE, Sim KB. Robust lip detection based on histogram of oriented gradient features and convolutional neural network under effects of light and background. *Optik*. 2017;136:462-469. doi:10.1016/j.ijleo.2017.02.017
34. Manda B, Dhayarkara S, Mitheran S, Viekash VK, Muthuganapathy R. "CADSketchNet" – an annotated sketch dataset for 3D CAD model retrieval with deep neural networks. *Comput Graph*. 2021;99:100-113. doi:10.1016/j.cag.2021.07.001

**How to cite this article:** Elen A. Covid-19 detection from radiographs by feature-reinforced ensemble learning. *Concurrency Computat Pract Exper*. 2022;34(23):e7179. doi: 10.1002/cpe.7179