IRBM 41 (2020) 212–222

Contents lists available at ScienceDirect

IRBM

www.elsevier.com/locate/irbm

Elsevier Masson France

EM|consulte

www.em-consulte.com

Original Article

A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models

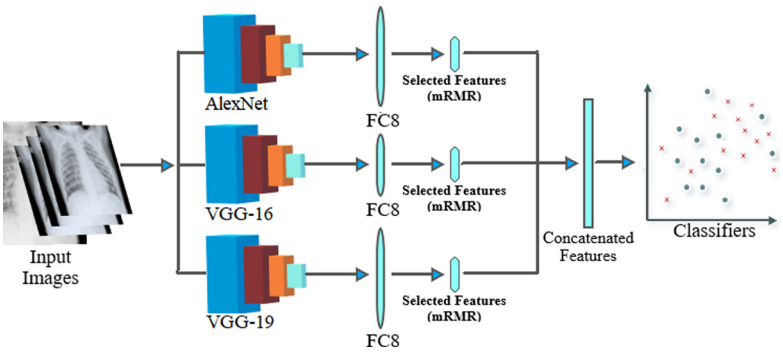
M. Toğaçar [a,*], B. Ergen [b], Z. Cömert [c], F. Özyurt [d]

[a] Department of Computer Technology, Fırat University, Elazığ, Turkey
[b] Department of Computer Engineering, Faculty of Engineering, Fırat University, Elazığ, Turkey
[c] Department of Software Engineering, Faculty of Engineering, Samsun University, Samsun, Turkey
[d] Department of Software Engineering, Faculty of Engineering, Fırat University, Elazığ, Turkey

HIGHLIGHTS

- This study applied deep learning models to detect pneumonia through the chest images.
- Image augmentation technique was applied to ensure balance between classes.
- We have extracted and combined features from the equivalent layers of CNN models.
- We have reduced combined features with the feature selection method (mRMR).
- We achieved the best result using various classifiers.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 May 2019
Received in revised form 17 October 2019
Accepted 25 October 2019
Available online 6 November 2019

Keywords:
Biomedical image processing
Diagnosis system
Pneumonia
Convolutional neural network
Feature selection

ABSTRACT

Pneumonia is one of the diseases that people may encounter in any period of their lives. Approximately 18% of infectious diseases are caused by pneumonia. This disease may result in death in the following stages. In order to diagnose pneumonia as a medical condition, lung X-ray images are routinely examined by the field experts in the clinical practice. In this study, lung X-ray images that are available for the diagnosis of pneumonia were used. The convolutional neural network was employed as feature extractor, and some of existing convolutional neural network models that are AlexNet, VGG-16 and VGG-19 were utilized so as to realize this specific task. Then, the number of deep features was reduced from 1000 to 100 by using the minimum redundancy maximum relevance algorithm for each deep model. Accordingly, we achieved 100 deep features from each deep model, and we combined these features so as to provide an efficient feature set consisting of totally 300 deep features. In this step of the experiment, this feature set was given as an input to the decision tree, *k*-nearest neighbors, linear discriminant analysis, linear regression, and support vector machine learning models. Finally, all models ensured promising results, especially linear discriminant analysis yielded the most efficient results with an accuracy of 99.41%. Consequently, the results point out that the deep features provided robust and consistent features for pneumonia detection, and minimum redundancy maximum relevance method was found a beneficial tool to reduce the dimension of the feature set.

* Corresponding author.
E-mail addresses: mtogacar@firat.edu.tr (M. Toğaçar), bergen@firat.edu.tr (B. Ergen), zcomert@samsun.edu.tr (Z. Cömert).

https://doi.org/10.1016/j.irbm.2019.10.006
1959-0318/

1. Introduction

Image processing is a technology that provides fast and accurate results on computer systems [1,2]. Recently, the model that is often preferred in image processing is the deep learning model. Convolution Neural Network (CNN), which is the sub-branch of deep learning, has a multi-layered structure in contrast to machine learning. Deep learning models have demonstrated their effectiveness in many areas [3–6]. One of its areas of interest is biomedical engineering. The introduction of deep learning models in the biomedical field has contributed to this area [7].

From past to present, infectious diseases are one of the most important factors that threaten human health. Pneumonia is one of the leading infectious diseases [8]. It is the inflammation caused by the virus and bacteria that microscopically adversely affect the air sacs [9]. Approximately 7% of the world's population is affected by pneumonia every year, and 4 million of the affected patients face fatal risks [10]. So, early diagnosis is important in such diseases. Typical symptoms of pneumonia include chest pain, shortness of breath, cough, etc. are located. The diagnosis of pneumonia in childhood is difficult due to the low sensitivity of microbiological tests and weak clinical finding [11]. Thus, the chest X-ray has become an important diagnostic tool in the diagnosis of pneumonia in children. Diagnostic tools include sputum culture and chest X-rays [12]. It is a time-consuming process in the medical area that physicians look and diagnose chest X-rays [13]. It is an extremely positive development in terms of time and cost to diagnose by utilizing existing technological means and software. The deep learning models have ensured more efficient results compared to traditional methods through Chest X-ray images obtained from pneumonia patients [14,15].

Computer-assisted diagnosis (CAD) systems have been proposed to improve accurate diagnostic performance and prevent possible errors in medical applications. In other words, CAD aims to recognize the occurrence of pneumonia cases and to prevent cases that may adversely affect public health [16]. In this issue, deep learning methods have been used to extract the features of the chest X-rays. This study suggests that the proposed CAD system reveal the most efficient features of pneumonia and contribute positively to the classification results.

Advances in deep learning have recently played an effective role in the biomedical field. Deep learning has pronounced its name with the high generalization performances for many difficult problems related to the field. It has yielded remarkable performance results compared to state-art of the models for many diseases [15]. Automatic analysis of chest X-rays images with CNN models has begun to gain further interest. Several procedures such as tuberculosis detection [17], segmentation [18], mass detection and classification [19] were performed on X-ray images. In addition, hybrid (combined) CNN models showed better results than CNNs. Detection of pneumonia was performed by selecting specific regions of interest (ROI) on X-ray images [14]. Transferring technique was also used by creating multiple label classes over X-ray images [20]. Chest X-ray and computed tomography images were used for chest segmentation [21].

In this study, pre-trained CNN models, which are AlexNet [22], VGG-16 and VGG-19 [23] with a similar layer structure were used to determine pneumonia. The dataset used in this study has two classes as shown in Fig. 1. Furthermore, the data augmentation method was also used to balance the distribution of samples over the classes. In the next step, each CNN model was separately applied to the dataset to extract the local discriminative features. The dimension of the obtained deep features were reduced using the minimum redundancy maximum relevance (mRMR) algorithm, and 100 features were selected for each CNN architecture [24–26]. Then the features obtained by mRMR feature selection algorithm were combined and this feature set was applied as the input to decision tree (DT) [27], *k*-nearest neighbors (*k*NN) [28], linear discriminant analysis (LDA) [29], linear regression (LR) [30], and support vector machine (SVM) [31]. In short, this study shows that the combination of the deep features obtained from the common layers of different CNN architectures increased the performance of the classification process. In addition, in the last step of the proposed approach, it can be observed that the consolidated deep features contributed to the classification performance with less but better features by using the mRMR feature selection method.

This study is briefly arranged as follows: In Section 2, short information about the used publicly available dataset is given. In Section 3, the existing CNN models, feature selection method, data augmentation method, machine learning method, optimization method, and the proposed method are presented briefly. The results are given in Section 4. The discussion is presented in Section 5. Lastly, the conclusion remarks are given in Section 6.

2. Dataset

The pneumonia dataset consists of a total of 5,849 samples, including 1,583 normal and 4,266 pneumonia images, respectively. The dataset was collected by taking the chest X-ray images of the volunteer patients.

The dataset consists of two classes and it is publicly accessible. The images were obtained from a total of 5,849 patients and 3,883 of these patients are children. Also, the images in the dataset are of RGB channels, eight-bit depth in each channel, have non-fixed (variable size) image size and images' format is JPEG [32].

3. Methods

3.1. Convolutional neural networks

CNNs are designed to automatically learn spatial hierarchies of features through a backpropagation algorithm. CNNs are designed using basic multiple building blocks such as convolution, pooling and fully connected (FC) layers [33–35]. The task of the convolutional layer is based on the process of circulating the input image with the selected filter. The size of the filter can be as 3×3, 5×5 or 7×7 pixels. Thus, it is created the input of the next layer ($m_2 \times m_3$) with the filter applied to the image. Activation maps occur as a result of this convolution process. Activation maps have local distinctive features [36]. Each convolutional layer has a filter (m_1). The output Y_i^l of layer l consists of $m_1^{(l)}$ feature maps of size $m_2^{(l)} \times m_3^{(l)}$. The ith feature map, showed $Y_i^{(l)}$, is calculated according to Eq. (1). Where $B_i^{(l)}$ represents the bias matrix and $K_{i,j}^{(l)}$ represents the filter size [37].

$$Y_i^{(l)} = f\left(B_i^{(l)} + \sum_{j=1}^{m_i^{(l-1)}} K_{i,j}^{(l)} \times Y_j^{(l-1)}\right) \tag{1}$$

The pooling layer keeps image features, reduces image size and costs and also keeps image information intact [38]. In addition, this structure decreases the number of parameters [39]. The pooling layer (l) has two parameters, the spatial size of the $F^{(l)}$ filter and the $S^{(l)}$ step. It receives input data in the size of $m_1^{(l-1)} \times m_2^{(l-1)} \times m_3^{(l-1)}$ and again provides an output volume of $m_1^{(l)} \times m_2^{(l)} \times m_3^{(l)}$. Briefly, the operation of the pooling layer is shown in Eqs. (2), (3) and (4).

$$m_1^{(l)} = m_1^{(l-1)} \tag{2}$$

$$m_2^{(l)} = (m_2^{(l-1)} - F^{(l)})/S^{(l)} + 1 \tag{3}$$

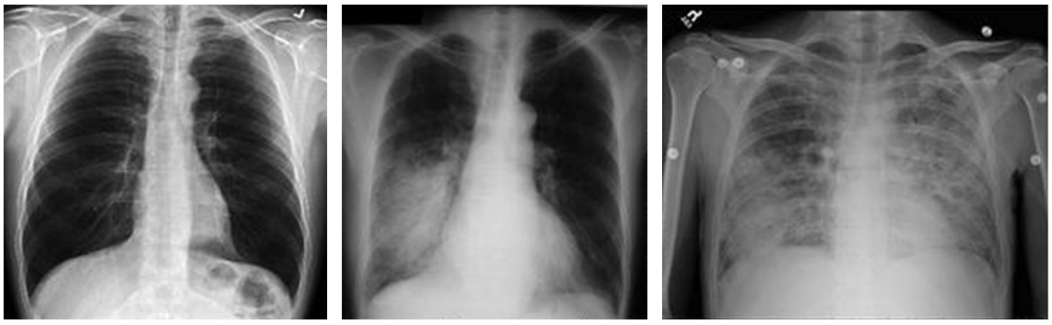


Fig. 1. (a) Example of a normal class, (b) and (c) examples of pneumonia class.

$$m_3^{(l)} = (m_3^{(l-1)} - F^{(l)})/S^{(l)} + 1 \quad (4)$$

The FC layers are respectively assigned to a single vector by flattening the features from the previous layer, they perform the update of the weights and give the last possible values for each label [34]. Fully interconnected layers in a convolutional network are practically a multilayer sensor for mapping $m_1^{(l-1)} \times m_2^{(l-1)} \times m_3^{(l-1)}$ [37]. The processing steps of the FC layer are shown in Eq. (5), if $(l-1)$ is an FC layer;

$$Y_i^{(l)} = f(Z_i^{(l)}) \quad \text{with } Z_i^{(l)} = \sum_{j=1}^{m_1^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)} \quad (5)$$

In this section, the AlexNet [22], VGG-16, and VGG-19 [23] models are considered and described briefly. These three models have already announced their names in ImageNet competitions. The AlexNet architecture consists of convolutional, pooling and fully connected layers. In the AlexNet, the input size of the image is set to 227×227 pixels. The size of the filter can be selected as 3×3 or 5×5 pixels in this model.

The VGG-16 also consists of convolution, pooling and fully connected layers, as in the AlexNet. It occurs of a total of 16 layers excluding pooling layers [23]. This architecture has an increasing network structure. In the VGG-16, the input size of the image is set to 224×224 pixels. Image filter size is selected as 3×3 pixels. The output layer of this architecture called Softmax consists of an activation function that gives the class probabilities [40].

VGG-19 consists of totally 19 layers, 16 of which are convolutional and 3 fully connected layers. Filters are used in convolutional layers to reduce the number of parameters in such deep networks. The size of the selected filter in this architecture is 3×3 pixels. The number of steps is two and the maximum pooling layer is used. The VGG-19 contains about 138 million computational parameters [41].

In this study, the last activation layers (FC8) have been used for deep feature extraction. The common point of these three architectures is that they have FC8 layers [42,43]. The architectural design of the AlexNet, VGG-16 and VGG-19 models is illustrated in Fig. 2.

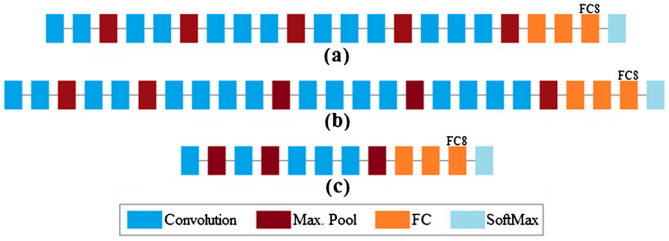


Fig. 2. The schematic diagram of (a) VGG-16, (b) VGG-19 and (c) AlexNet models.

Fig. 3. An original sample from the normal class.

3.2. Image augmentation techniques

Data augmentation techniques are used together with traditional methods or deep learning methods to improve the accuracy of classification [44,45]. In this study, the image augmentation algorithm was used by using the Keras library in Python [46]. The rotation, width and height change, cutting, zooming, horizontal turning, brightness, and filling operations were performed for normal class images. The image rotation degree was set to be randomly generated from 0 to 20. Width-height change, zoom and horizontal rotation ratio were selected as 0.15. An example of a normal image is showed in Fig. 3. The results of image augmentation algorithm performed on this image are showed in Fig. 4.

In this study, the image augmentation techniques were applied only to normal images in order to balance the distribution of the samples over the classes. The number of normal samples in the dataset was increased from 1,583 to 4,266 by performing the image augmentation techniques. In this manner, the number of samples for each class was equalized. This equal distribution makes it possible to use all of the data instead of selecting random data during the training process. It is expected that this situation increases the accuracy of the training and positively affects the classification results.

3.3. Machine learning method

Decision Tree (DT) is usually one of the methods used in data mining. It uses the methods of knowledge acquisition and entropy. Another characteristic feature of DT is the ability to classify data sets based on the graph-based rules. The features derived from the dataset are combined to perform the process of rules generated from the DT [47].

k-nearest neighbors (kNN) is a nonparametric method based on distance measurement between the features of the training set [48]. The classification is realized using the distance between the selected features and the k-nearest neighbors. The Euclidean

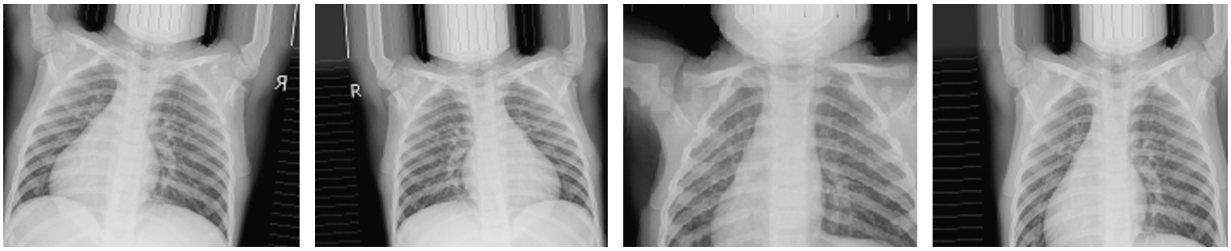


Fig. 4. A subset of new images obtained from an original normal image by using the image augmentation technique.

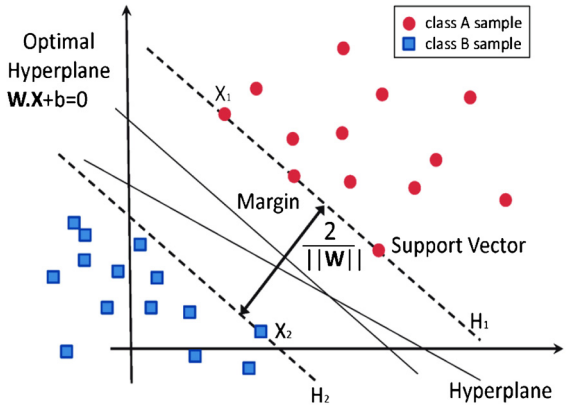


Fig. 5. Linear SVM classification process [55].

formula given in Eq. (6) can be used to determine the distances among the features. In addition, different distance functions, such as Cosine, Euclidean or Manhattan distance, can also be used for this purpose [49]. In this study, k value was adjusted to three and the Euclidean was used as a distance function.

$$d_{(i,j)} = \sqrt{\sum_{k=1}^{p} \left(X_{ik} - X_{jk}\right)^2} \tag{6}$$

Linear discriminant analysis (LDA) is a commonly used linear method for dimension reduction in the preprocessing step of the analyzes [50]. In the classification task, it is preferred to avoid excessive time loss and reduce the cost of calculation [51].

Linear regression (LR) is one of the most basic machine learning techniques. It aims to obtain a function by linking the input parameters to the target values. The relationship to be established is called cause and effect. After capturing a linear relationship between cause and effect, it goes into the regression or classification processes. It performs the regression or classification tasks by estimating the target values according to the characteristics of the data [52].

Support vector machine (SVM) is a supervised machine learning method that can be used for regression and classification problems. SVM searches a hyperplane that separates the positive data from the negative data with maximum margin [53,54]. The training of a SVM classification process is showed in Fig. 5. The formula for the output of a linear SVM is given in Eq. (7), where $\vec{w}$ is the normal vector to the hyperplane, and $\vec{x}$is the input vector.

Maximizing margins can be defined as an optimization problem: minimize Eq. (8) subject to Eq. (9) where x_i is ith training example and y_i is the correct output of the SVM for ith training example [56].

$$u = \vec{w} \cdot \vec{x} - b \tag{7}$$

$$\frac{1}{2}\left\|\vec{w}\right\|^2 \tag{8}$$

$$y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1, \forall i \tag{9}$$

Table 1
Feature selection criteria of mRMR algorithm.

Feature criteria	Equations
Mutual Information Difference (MID)	$max_{F_i \epsilon \Omega_S}\left[I(F_i, H) - \frac{1}{\|S\|}\sum_{F_j \epsilon S} I(F_i, F_j)\right]$
Mutual Information Quotient (MIQ)	$max_{F_i \epsilon \Omega_S}\left[I(F_i, H)/\frac{1}{\|S\|}\sum_{F_j \epsilon S} I(F_i, F_j)\right]$

Finally, the Softmax method used in this study is a generalized form of the LR method. In other words, it is used in classification processes where the number of classification label is two or more [57].

3.4. Feature selection

The main objective of the mRMR method is to select the best features and to dilute the ones that are low [58–60]. This method handles each feature separately from the dataset and uses the mutual information between them to measure the level of similarity between the two features, M and N, using the $I(M, N)$:

$$I(M, N) = \sum_{n \in N}\sum_{m \in M}(m, n)\log\left(\frac{p(m, n)}{p_1(m)p_2(n)}\right) \tag{10}$$

Herein, M and N combined probability is defined as the distribution function. $P_1(m)$ and $P_2(n)$ respectively show the marginal probability distribution functions of M and N random variables. Each property of the equation is defined as f_i. Each property is defined in the vector created in K-size ($f_i = [f_i^1, f_i^2, f_i^3, \ldots, f_i^K]$). The information formed between the variables i and j is defined by $I(F_i, F_j)$. The variables i and j are valued from 1 to d, where d denotes the size of the vector. The similarity between any property and the class label is represented by the H symbol. The S symbol indicates the number of features of the selected set. In order to understand that the best features are selected, the following Eq. (11) and (12) must be satisfied:

$$minW, W = \frac{1}{|S|^2}\sum_{F_i F_j \in S} I(F_i, F_j) \tag{11}$$

and the other is the maximum relevance condition:

$$maxV, V = \frac{1}{|S|}\sum_{F_i \in S} I(F_i, H) \tag{12}$$

The combined equations that form two selection criteria for the mRMR method are shown in Table 1.

Finally, extraction of features with mRMR algorithm was performed using Python [60]. The mRMR method was applied to the features provided the CNNs. 1000 features of each architecture were applied as the input to the mRMR algorithm and the number of this feature set was reduced to 100 features. In the next step, this feature set were combined.

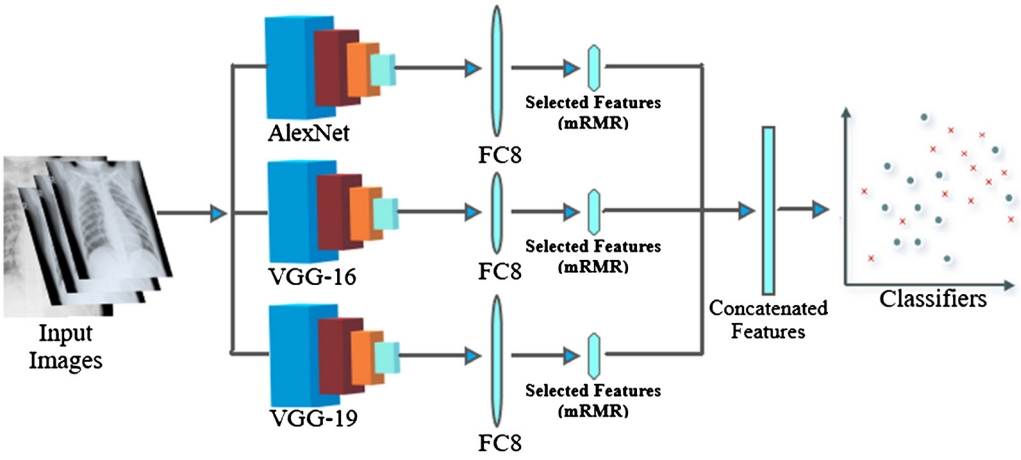


Fig. 6. Demonstration of the proposed approach.

Table 2
Parameter values of the present proposed approach used in this study.

Used software	CNN architecture	Image size	Optimization	Momentum	Decay	Beta	Mini batch	Learning rate
MATLAB	AlexNet	227×227	SGD	0.9	1e-6	–	16	0.0001
	VGG-16	224×224						
	VGG-19	224×224						

3.5. Optimization method

The main purpose of optimization methods is to update the weights at every stage until the best learning in CNN is realized. Each method performs an update process. In the Stochastic Gradient Descent (SGD) method, the weights update is performed for each training set. Because of this reason, it tries to achieve the goal as early as possible time [61,62]. The formulation of SGD optimization is shown in Eq. (13). Here, Θ is the weight vector to be updated, α is the learning coefficient and $\nabla_{\Theta} J(\Theta)$ is the cost function.

$$\Theta_t = \Theta_{t-1} - -\alpha \nabla_{\Theta} J\left(\Theta; x^i, y^i\right) \tag{13}$$

3.6. Proposed CNN method

In the proposed approach, the default parameter values were used in the three CNNs. The input data was used separately for each CNN. 1000 features were provided by each CNN. The mRMR feature selection method was used to select the best features out of every 1000 features. In the last stage of the proposed method, the best performance was achieved by using different classifiers.

The proposed method is shown in Fig. 6. The proposed method consists of three stages. In the first stage, the number of samples in the normal class is less than the number of samples in the pneumatic class. The aim is to equalize the number of samples for each class. Thus, all samples can be used in a balanced manner. The CNNs were used without pre-training. Only the FC8 layer was used. The FC8 layer ensures totally 1000 features. In the second stage, with the mRMR algorithm, the most relevant 100 features were selected from each CNN. Afterward, classification results of DT, *k*NN, LDA, LR and SVM classifiers were compared. In the third stage, the selected 100 features from each model were consolidated. Several combinations were tried and 200 features were obtained considering only two CNNs. Also, 300 features were obtained using the three models. Then, the classification performances of five machine learning models were compared.

4. Results

In this study, MATLAB (R2018b) and Python were utilized to realize the experiment. Pneumonia images were divided into two parts with rates as 70% training set and 30% test set, respectively. The pretrained deep models were frequently adopted to a new task through transfer learning approach. The values of the training parameters of pretrained CNNs used in the experiment are given in Table 2. As inferred from Table 2, these parameters were used with their default values. Also, the mini-batch size was adjusted to 16. Mini-batch contributes to learning processes by balancing the convergence rate of the network as well as accurate estimation [63]. However, the size of the mini-batch was not increased more since this is insufficient in terms of time-consumption and memory usage. The reason for reducing the mini-batch size suffered from the insufficient graphics card in computer hardware. The training graphs of the pre-trained CNN models are given in Fig. 7.

As seen in Fig. 7, the training processes of the models were completed with high accuracy. However, the validations processes of the models show that the models could not completely convergent. The oscillations in the validation loss point out that the pre-trained CNN models could not entirely learn the local discriminative features of pneumonia.

The cross-validation was not used in the proposed approach. Instead of the cross-validation method, the holdout validation technique was used since the dataset ensures enough samples for training and test sets. The image augmentation was only performed over the normal class to balance the distribution of the samples over the classes.

The models were compiled with graphics processing unit (GPU) support. The simulation environment was run on 64-bit Windows 10 operating system. Other hardware details of the computer are that NVIDIA GeForce 2 GB graphics card, Intel© i5-Core @2.5 GHz processor and 8 GB RAM.

In order to measure the performances of the models, Accuracy (Acc), sensitivity (Se), and specificity (Sp), precision (Pr), and F-score metrics derived from confusion matrix were used and the formulations of the metrics are described as follows [64]:

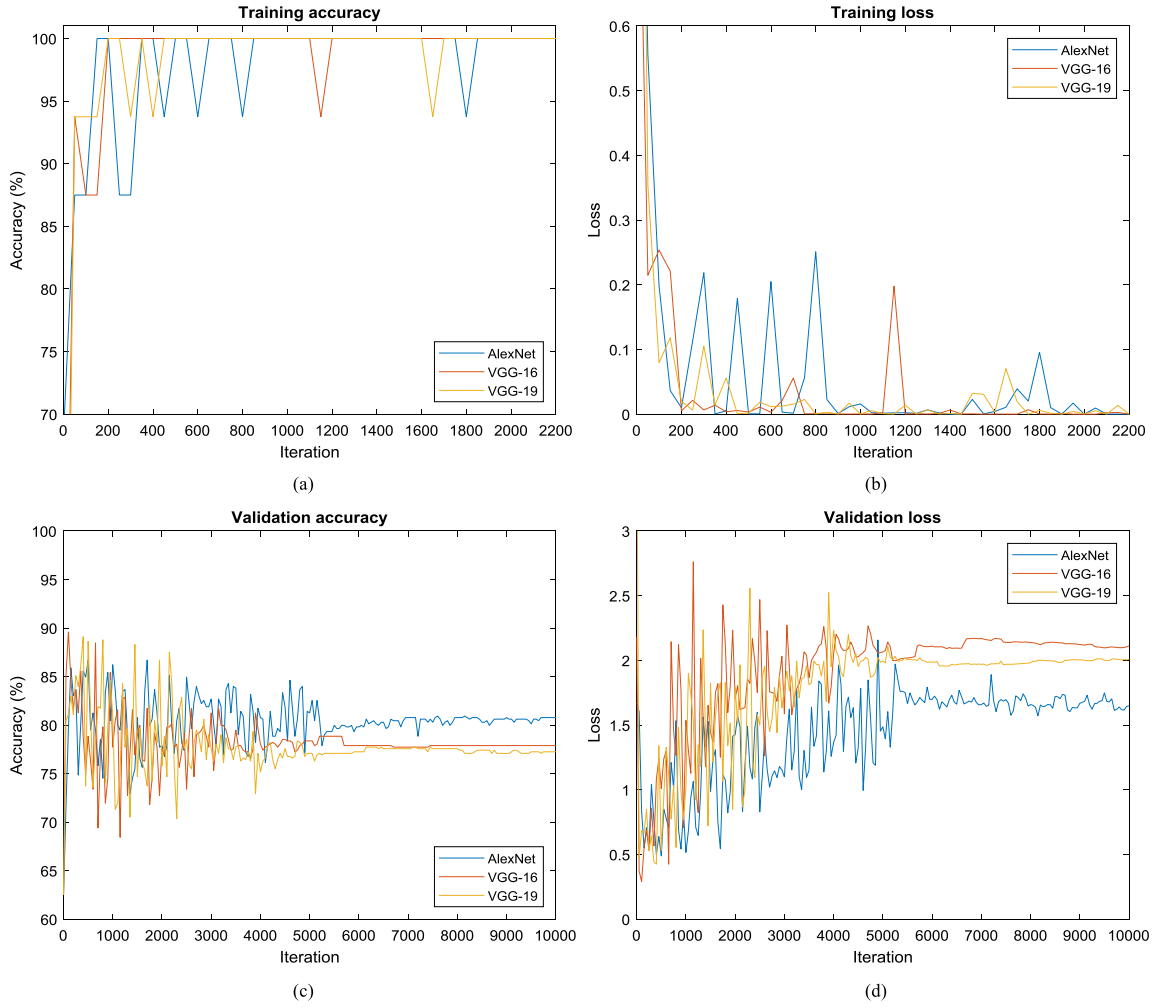


Fig. 7. The training graphs of the pre-trained CNN models. (a) Training accuracy, (b) Training loss, (c) Validation accuracy, (d) Validation loss.

$$Acc = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FN}) + (\mathrm{FP} + \mathrm{TN})} \tag{14}$$

$$Se = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FN})} \tag{15}$$

$$Sp = \frac{(\mathrm{TN})}{(\mathrm{TN} + \mathrm{FP})} \tag{16}$$

$$Pr = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FP})} \tag{17}$$

$$F - Score = \frac{(2\mathrm{xTP})}{(2\mathrm{xTP} + \mathrm{FP} + \mathrm{FN})} \tag{18}$$

where true positive (TP) represents the number of pneumonia images classified as diseased pneumonia whereas true negative (TN) represents the number of normal images classified as normal. Also, false positive (FP) represents the number of normal images incorrectly classified as diseased pneumonia while false negative (FN) represents the number of diseased pneumonia images misclassified as normal.

The experiment carried out in this study consists of four steps. In the first step, the original dataset was used without applying any image augmentation techniques. An end-to-end training approach was adopted to classify the pneumonia images. The classification results of the first step are given in Table 3. AlexNet, VGG-16 and VGG-19 models were used to extract features. The best result was obtained by using the SVM classifier fed with the features provided by VGG-19. The success rate of SVM was 94.84%.

In the second step, as mentioned before, the image augmentation methods were used to equalize the number of samples over the classes. The dataset was divided into two parts as training and test with 70% and 30% rates, as in the first step. In this step, the holdout validation was also applied. The augmented dataset was processed by three models. 1000 features from the FC8 layer of each model were extracted. The results of the five machine learning models are shown in Table 4. The best result was obtained by the LDA classifier that was fed with the features provided by the AlexNet. The success rate was 96.83%. Here, the number of the normal class with the augmentation technique was balanced by the number of the pneumatic class. Fig. 8 shows that the augmentation method contributed to improving classification success, except DT.

In the third step, the most relevant 100 features selected by the mRMR method were considered. The augmented dataset was used in this step setup. 30% of the samples were processed as test data and a holdout validation method was used. It is shown in Table 5 that the mRMR method gives better results with fewer features in the classification of each model according to the results in Table 3 and Table 4. The best results were ensured with the combination of three deep CNN models. The best success was achieved by the selected 300 features. The most efficient success rate was 99.41% that was yielded by LDA classifier. The results are reported in Table 5. Receive optical characteristic curve (ROC) and confusion matrix of LDA classifier are shown in Fig. 9.

Table 3
The results obtained by applying CNN models in combination with five classifiers to original pneumonia data.

CNN model	Classifier	# of features (FC8)	Acc	Se (%)	Sp (%)	Pr (%)	F-score (%)
AlexNet	DT	1000	85.99	84.17	88.0	88.61	86.33
	*k*NN		90.09	85.45	96.13	96.62	90.69
	LDA		**93.57**	92.58	94.61	94.73	93.64
	LR		92.31	91.34	93.32	93.46	92.39
	SVM		92.73	91.08	94.52	94.73	92.86
VGG-16	DT	1000	88.41	89.74	87.17	86.71	88.20
	*k*NN		91.46	88.76	94.57	94.94	91.74
	LDA		92.83	92.29	93.39	93.46	92.87
	LR		91.57	90.87	92.29	92.41	92.63
	SVM		**94.73**	95.30	94.18	94.09	94.69
VGG-19	DT	1000	89.36	88.45	90.30	90.51	89.47
	*k*NN		93.47	90.55	96.83	97.05	93.69
	LDA		92.83	92.29	92.39	93.46	92.87
	LR		91.99	90.78	93.28	93.46	92.10
	SVM		**94.84**	94.18	95.51	95.57	94.87

Table 4
The classification results obtained by applying CNN models in combination with five classifiers to the augmented dataset.

CNN model	Classifier	# of features (FC8)	Acc (%)	Se (%)	Sp (%)	Pr (%)	F-score (%)
AlexNet	DT	1000	87.03	85.52	88.66	89.13	87.29
	*k*NN		93.24	92.28	94.24	94.38	93.32
	LDA		**96.83**	96.80	96.87	96.88	96.84
	LR		96.29	96.33	96.25	96.25	96.29
	SVM		96.52	96.85	96.20	96.17	96.51
VGG-16	DT	1000	86.71	85.82	87.65	87.97	86.88
	*k*NN		93.79	94.10	93.48	93.44	93.77
	LDA		**96.68**	96.79	96.57	96.56	96.68
	LR		95.04	95.57	94.51	94.45	95.01
	SVM		96.33	96.62	96.04	96.02	96.32
VGG-19	DT	1000	88.71	88.73	88.68	88.66	88.70
	*k*NN		94.26	94.92	93.61	93.51	94.21
	LDA		**96.56**	96.41	96.71	96.72	96.57
	LR		95.47	95.68	95.26	95.23	95.45
	SVM		96.48	96.19	96.78	96.79	96.49

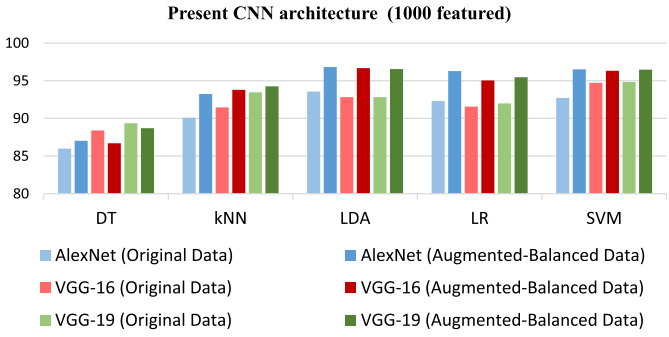


Fig. 8. Comparison of the results obtained in the original dataset of present CNN models with the results of the was balanced dataset on a class basis.

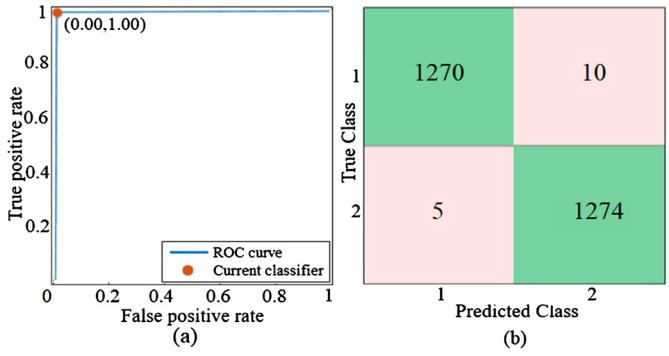


Fig. 9. (a) The ROC curve of LDA classifier that was fed with 300 features selected by mRMR, (b) The confusion matrix of LDA classifier that was fed with 300 features selected by mRMR.

The linear graph of the classification results of the models in Table 4 and the linear graph of the classification results in Table 5, are showed in Fig. 10. The efficacy of the mRMR method was also observed in the analysis results of Fig. 10. Fig. 10 shows that it yields better results with 100 features received from three models.

In the fourth step, the non-augmented dataset was used. The proposed approach was performed using the original dataset. 30% of the dataset was allocated as test data. Using the mRMR method, the best 100 features obtained from CNNs were combined and a new feature set with 300 features was generated. In the next process, classification was performed. The results of the classification analysis are shown in Table 6. The best classification success was achieved by SVM at 98.21%. The ROC curve and confusion matrix of this experiment is shown in Fig. 11.

5. Discussion

Pneumonia is one of the common diseases among children. If this disease is not diagnosed in a timely manner, the probability of death may fairly high [65]. The early diagnosis of this disease is associated with rapid and accurate results of the image processing techniques as regarding the computation approaches. In this field, CNN models have a great advantage in terms of giving faster and better results compared to traditional machine learning methods [66]. Especially, these models have a powerful ability to extract the discriminative local features. Furthermore, the models can also reflect this ability to the classification process. As a result, numerous studies have been focused on machine learning and deep learning models to early diagnose pneumonia disease through pneumonia data. With the proposed method, we carried out a pneumonia clas-

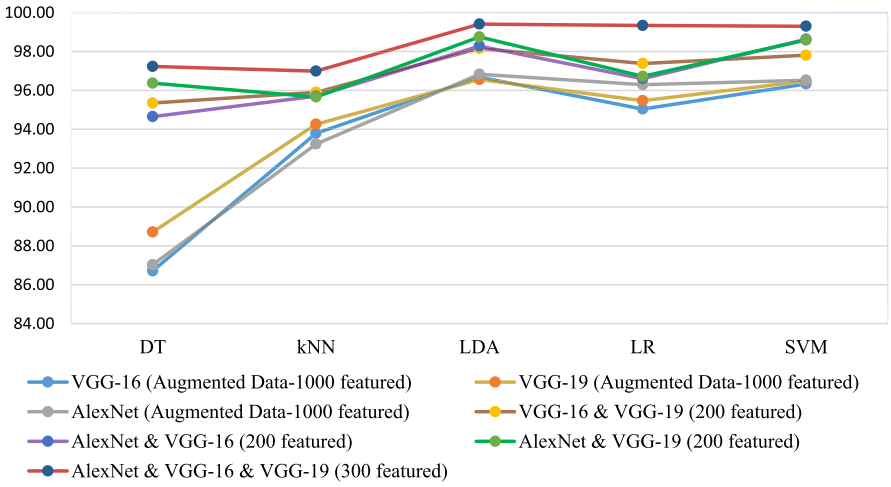


Fig. 10. Comparison of classification results obtained by data augmentation technique and classification results obtained by reducing the feature with mRMR method.

Table 5
The classification results obtained by applying the mRMR method to the features obtained from CNN models applied to augmented pneumonia data.

CNN models	Classifier	# of features (mRMR)	Acc (%)	Se (%)	Sp (%)	Pr (%)	F-score (%)
AlexNet & VGG-16	DT	(100+100) 200	94.65	93.65	95.68	95.78	94.70
	*k*NN		95.70	96.35	95.07	95.0	95.67
	LDA		98.28	98.13	98.43	98.44	98.28
	LR		96.60	97.23	95.99	95.93	96.58
	SVM		**98.63**	98.75	98.52	98.51	98.63
AlexNet & VGG-19	DT	(100+100) 200	96.37	95.90	96.84	96.88	96.39
	*k*NN		95.66	96.95	94.44	94.30	95.60
	LDA		**98.75**	98.60	98.90	98.91	98.75
	LR		96.72	96.65	96.79	96.80	96.72
	SVM		98.59	98.52	98.67	98.67	98.59
VGG-16 & VGG-19	DT	(100+100) 200	95.35	95.10	95.60	95.63	95.36
	*k*NN		95.90	96.59	95.22	95.16	95.87
	LDA		**98.16**	98.51	97.83	97.81	98.16
	LR		97.38	98.10	96.69	96.64	97.36
	SVM		97.81	98.26	97.37	97.34	97.80
AlexNet & VGG-16 & VGG-19	DT	(100+100+100) 300	97.23	96.82	97.63	97.66	97.24
	*k*NN		96.99	98.63	95.46	95.31	96.94
	LDA		**99.41**	99.61	99.22	99.22	99.41
	LR		99.34	99.61	99.07	99.06	99.33
	SVM		99.30	99.37	99.22	99.22	99.30

Table 6
Classification results obtained by applying the proposed approach to augmented pneumonia data and non-augmented pneumonia data.

CNN models / dataset status	Classifier	# of features (mRMR)	Acc (%)	Se (%)	Sp (%)	Pr (%)	F-score (%)
AlexNet & VGG-16 & VGG-19	DT	(100+100+100) 300	89.04	89.19	88.89	88.82	89.01
/	kNN		96.31	94.34	98.46	98.52	96.39
original dataset	LDA		98.10	98.93	97.31	97.26	98.09
(non-augmented dataset)	LR		98.0	98.51	97.50	97.47	97.99
	SVM		**98.21**	98.51	97.91	97.89	98.20
AlexNet & VGG-16 & VGG-19	DT	(100+100+100) 300	97.23	96.82	97.63	97.66	97.24
/	*k*NN		96.99	98.63	95.46	95.31	96.94
balanced dataset	**LDA**		**99.41**	99.61	99.22	99.22	99.41
(augmented dataset)	LR		99.34	99.61	99.07	99.06	99.33
	SVM		99.30	99.37	99.22	99.22	99.30

sification. The results of the study were compared with the related studies as shown in Table 7.

Aditya Sriram et al. proposed a new neural network architecture that uses radon projections to both classify and represent medical images. The name of the model was Projectron. They used radon transfer technique in their studies. The used Radon transform in the study was an established technique that can reconstruct images from parallel projections. Aditya Sriram et al. used a global radon transformation with angles at equal angles to each image. The goal was to facilitate the linear separation of transformation outputs. They used 30% of the data as a test set in the study. Multi-Layer Perceptron (MLP) was selected as the classifier and achieved 70.03% classification success.

A.A. Saraiva et al. designed a CNN model with 300 × 300 data inputs. The CNN model consisted of seven convolutional, three maximum pooling and three dense layers. They used the ADAM optimization method in the model and employed the Softmax method to get the probabilities of the classes. The 5-fold cross-

Table 7
Comparison of the success of studies using the same dataset with the success of the proposed approach.

Study	Year	Model	Method or optimization	Classifier	Acc (%)
Aditya Sriram et al. [67]	2019	Projectron	Radon transform	Multi-Layer Perceptron (MLP)	70.03
A.A. Saraiva et al. [68]	2019	The Designed CNN	ADAM	Softmax	95.30
Ioannis Livieris et al. [69,70]	2019	Machine Learning	Semi-supervised learning (SSL)	Self-labeled algorithm	83.49
Karan Jakhar et al. [71]	2018	Deep CNN	–	Softmax	84.0
Daniel S. Kermany et al. [72]	2018	Inception V3	Transfer Learning & ADAM	Softmax	92.80
Proposed method	2019	CNNs	Without data augmentation, consolidated deep features, mRMR & SGD	SVM	**98.21**
Proposed method	2019	CNNs	Consolidated deep features, mRMR & SGD	LDA	**99.41**

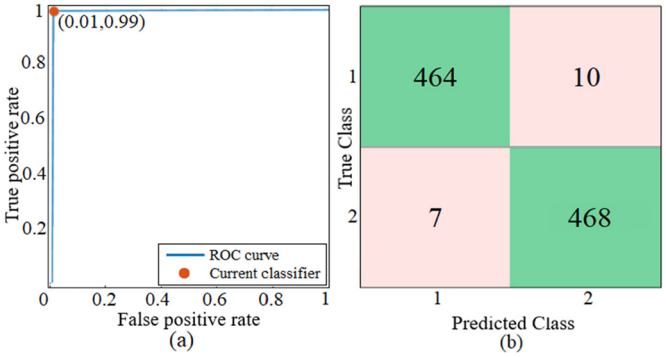


Fig. 11. Applying the proposed approach to the non-augmented dataset. (a) The ROC curve of SVM classifier that was fed with 300 features selected by mRMR, (b) The confusion matrix of SVM classifier that was fed with 300 features selected by mRMR.

validation method was also utilized. As a result, the classification success was 95.30%. It has been observed that the optimization method was chosen by A.A. Saraiva et al. In addition to the SGD method we use in our study, it is thought that using other optimization methods will have a positive effect on the classification results. In addition, the resolution of the input image in their study (300×300) contributed positively to the feature obtained.

Ioannis Livieris et al. used "self-labeled algorithm" for X-ray's classification based on a maximum-probability voting scheme. They categorized each image by the algorithms (number of algorithms: 3). At a later stage, more than one classified label of an image was formed. They classified each image according to the absolute majority class. The three algorithms they used are semi-supervised learning algorithms. They achieved a classification success of 83.49% using pneumonia dataset. The reason they are not effective in the tagging event in this study is that the number of normal class data is not equal to the number of pneumonia data. We recommend the method of augmenting data to Ioannis Livieris et al. In this case, since the number of labels of the normal class data will increase, the classification performance will increase directly.

Karan Jakhar et al. used the deep CNN model with 10 fold cross-validation techniques. They used various classifiers (SVM, LR, Random Forest, etc.) with the DCNN model and performed the classification process. The best achievement was obtained by using the DCNN model. Their success rate was 84%. It is an advantage to find the best classifier method to use various classifiers in their study. We used different classifiers in our study. Thus, we chose the best among the classifiers. However, they can use different optimization methods and parameters in the DCNN model they offer to improve the model.

Daniel S. Kermany et al. used a transfer learning method included some of the data of traditional approaches in their architecture and trained a neural network. They applied a CNN model using Keras Tensorflow library and performed the classification of pneumonia using the transfer learning method with a much faster, less training sample. They carried out the transfer learning process using the Inception V3 model. They used the holdout method for the validation process. In the proposed model, the learning rate was 0.001, and the number of the epoch was set to 100. They used the SGD method for the training of the model, and the ADAM method for optimization. Thus, the achieved success was 92.80%. Daniel S. Kermany et al. used CNN models with transfer learning in their study. In the same way, we used CNN models with transfer learning in our study. The fact that our study gives better results than their study is related to the approach model we follow. That is, the combination of the layers of the CNN models and the effectiveness of the feature selection method contributed to the classification result.

In another study in Table 7 using the same dataset, normal and pneumonia data were not equilibrated. This situation reduced the success of the classification of other studies. In order to create a healthier feature in our own study, we balanced the data of normal class with the number of pneumatic data. In addition, by using the feature selection method in our study, we chose fewer features (efficient features) and obtained better yields than other studies. Thus, compared to the other studies shown in Table 7, we achieved superior classification success with more efficient time usage and low cost. By the way, we also realized fourth step in which the data augmentation techniques were not used and the deep features were consolidated. In this step, the most efficient results were provided by SVM classifier with 98.21% classification accuracy.

In this study, mRMR feature selection method was found as rather effective. The image augmentation techniques were used so as to provide a balanced dataset. The 1000 features obtained from the last fully connected layers were then reduced to 100 for each CNN model with the mRMR method. The most relevant features selected by mRMR were consolidated. The best efficient results were obtained by combining the deep features extracted from AlexNet, VGG-16, VGG-19 models and 99.41% success rate was achieved.

6. Conclusion

The current study focused on improving the classification accuracy of the pneumonia. The dataset consisting of X-ray images was separated into two different classes as normal or pneumonia. In this study, we utilized existing AlexNet, VGG-16, and VGG-19 CNN models as a feature extractor. For this specific task, we used the last fully-connected layer of the CNN models, and these deep features were used to feed DT, *k*NN, LDA, LR, and SVM machine learning models. The classification task was carried out without applying any preprocessing procedure to pneumonia images. To balance the distribution of the samples over the classes, the image augmentation techniques were applied to only normal classes. Then, the most relevant features were selected by applying the mRMR method. The most efficient results were obtained by combining all features provided by mRMR. Finally, the combination of LDA and mRMR methods provided the best performance with an accuracy of 99.41%, sensitivity of 99.61% and specificity of 99.22%

on the pneumonia dataset. The results of this study point out that the deep features provided robust and consistent features for pneumonia detection, and the mRMR method increased the efficiency of the classification. The classification performance of this proposed approach is superior to the previously attempted techniques on the same dataset.

In the future, we will investigate the performance of the proposed approach on different datasets. The proposed method can be generalized to design high-performance computer-aided diagnosis systems for other medical imaging tasks.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

References

[1] Toğaçar M, Ergen B. Deep learning approach for classification of breast cancer. In: 2018 Int Conf Artif Intell Data Process; 2018. p. 1–5.
[2] Peng EB, Zhang GT. Image processing technology research of on-line thread processing. Adv Mater Res 2014;908:555–8. https://doi.org/10.4028/www.scientific.net/amr.908.555.
[3] Del Fiol G, Michelson M, Iorio A, Cotoi C, Haynes RB. A deep learning method to automatically identify reports of scientifically rigorous clinical research from the biomedical literature: comparative analytic study. J Med Internet Res 2018;20:e10281. https://doi.org/10.2196/10281.
[4] Cömert Z, Fetal Kocamaz AF. Hypoxia detection based on deep convolutional neural network with transfer learning approach. In: Silhavy R, editor. Softw Eng Algorithms Intell Syst. Cham: Springer International Publishing; 2019. p. 239–48.
[5] Budak Ü, Cömert Z, Çıbuk M, Şengür A. DCCMED-Net: densely connected and concatenated multi encoder-decoder CNNs for retinal vessel extraction from fundus images. Med Hypotheses 2020;134:109426. https://doi.org/10.1016/j.mehy.2019.109426.
[6] Budak Ü, Cömert Z, Rashid ZN, Şengür A, Çıbuk M. Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. Appl Soft Comput 2019;85:105765. https://doi.org/10.1016/j.asoc.2019.105765.
[7] Bakator M, Radosav D. Deep learning and medical diagnosis: a review of literature. Multimodal Technol Interact 2018;2:47. https://doi.org/10.3390/mti2030047.
[8] Akter S, Shamsuzzaman SM, Jahan F. Community acquired bacterial pneumonia: aetiology, laboratory detection and antibiotic susceptibility pattern. Malays J Pathol 2014;36:97–103.
[9] Kolditz M, Ewig S. Community-acquired pneumonia in adults. Dtsch Arztebl Int 2017;114:838–48. https://doi.org/10.3238/arztebl.2017.0838.
[10] Shen Y, Tian Z, Lu D, Huang J, Zhang Z, Li X, et al. Impact of pneumonia and lung cancer on mortality of women with hypertension. Sci Rep 2016;6:20. https://doi.org/10.1038/s41598-016-0023-2.
[11] Oliveira LLG, Silva SA, Ribeiro LHV, de Oliveira RM, Coelho CJ, Ana Lúcia ALS. Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. Int J Med Inform 2008;77:555–64. https://doi.org/10.1016/j.ijmedinf.2007.10.010.
[12] Pletz MW, Rohde GG, Welte T, Kolditz M, Ott S. Advances in the prevention, management, and treatment of community-acquired pneumonia. F1000Res 2016;5. https://doi.org/10.12688/f1000research.7657.1. F1000 Faculty Rev-300.
[13] Lavine M. The early clinical X-ray in the united states: patient experiences and public perceptions. J Hist Med Allied Sci 2011;67:587–625. https://doi.org/10.1093/jhmas/jrr047.
[14] Rajaraman S, Thoma G, Antani S, Candemir S. Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. https://doi.org/10.1117/12.2512752, 2019.
[15] Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET / CT images. EJNMMI Res 2017. https://doi.org/10.1186/s13550-017-0260-9.
[16] Qin C, Yao D, Shi Y, Song Z. Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online 2018;17:113. https://doi.org/10.1186/s12938-018-0544-y.
[17] Stirenko S, Kochura Y, Alienin O, Rokovyi O, Gordienko Y, Gang P, et al. Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In: 2018 IEEE 38th Int Conf Electron Nanotechnol; 2018. p. 422–8.
[18] Gang P, Zhen W, Zeng W, Gordienko Y, Kochura Y, Alienin O, et al. Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer. In: 2018 Tenth Int Conf Adv Comput Intell; 2018. p. 878–83.
[19] Abiyev RH, Ma'aitah MKS. Deep convolutional neural networks for chest diseases detection. J Healthc Eng 2018;2018:4168538. https://doi.org/10.1155/2018/4168538.
[20] Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of deep learning approaches for multi-label chest X-ray classification. Sci Rep 2019;9:6381. https://doi.org/10.1038/s41598-019-42294-8.
[21] Mansoor A, Bagci U, Foster B, Xu Z, Papadakis GZ, Folio LR, et al. Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. Radiographics 2015;35:1056–76. https://doi.org/10.1148/rg.2015140232.
[22] Krizhevsky A, Sutskever I, Hinton GE. Machine learning and computer vision group deep learning with tensorflow; 2012.
[23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
[24] Che J, Yang Y, Li L, Bai X, Zhang S, Deng C. Maximum relevance minimum common redundancy feature selection for nonlinear data. Inf Sci (NY) 2017;409–410:68–86. https://doi.org/10.1016/j.ins.2017.05.013.
[25] Vaswani A, Noam S, Parmar N. MCRMR: maximum coverage and relevancy with minimal redundancy based multi-document summarization. Expert Syst Appl 2019;120:43–56. https://doi.org/10.1016/j.eswa.2018.11.022.
[26] Cıbuk M, Budak U, Guo Y, Ince MC, Sengur A. Efficient deep features selections and classification for flower species recognition. Measurement 2019;137:7–13. https://doi.org/10.1016/j.measurement.2019.01.041.
[27] Webb AR. Statistical pattern recognition. John Wiley & Sons; 2003.
[28] Akbulut Y, Sengur A, Guo Y, Smarandache F. NS-k-NN: neutrosophic set-based k-nearest neighbors classifier. Symmetry (Basel) 2017;9.
[29] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer; 2009.
[30] Hoffman JIE. Linear regression. In: Hoffman JIE, editor. Basic Biostat Med Biomed Pract. second ed. Academic Press; 2019. p. 445–89.
[31] Frunza M-C. Support vector machines. In: Frunza M-C, editor. Solving Mod Crime Financ Mark. Academic Press; 2016. p. 205–15.
[32] Mooney P. Chest X-ray images (pneumonia). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia, 2018.
[33] Zhang M, Li L, Wang H, Liu Y, Qin H, Zhao W. Optimized compression for implementing convolutional neural networks on FPGA. Electronics 2019;8:295. https://doi.org/10.3390/electronics8030295.
[34] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018;9:611–29. https://doi.org/10.1007/s13244-018-0639-9.
[35] Togacar M, Ergen B, Subclass Sertkaya ME. Separation of white blood cell images using convolutional neural network models. Elektron Elektrotech 2019;25:63–8. https://doi.org/10.5755/j01.eie.25.5.24358.
[36] Koushik J. Understanding convolutional neural networks. https://doi.org/10.1016/j.jvcir.2016.11.003, 2016.
[37] Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition; 2010. p. 92–101.
[38] O'Shea K, Nash R. An introduction to convolutional neural networks; 2015.
[39] Passalis N, Learning Tefas A. Bag-of-features pooling for deep convolutional neural networks. Proc IEEE Int Conf Comput Vis 2017:5766–74. https://doi.org/10.1109/iccv.2017.614.
[40] Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: comparison of trends in practice and research for deep learning. arXiv:1811.03378, 2018.
[41] Huang Z, Nasrullah Wen J, Song S, Mateen M. Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry (Basel) 2018;11:1. https://doi.org/10.3390/sym11010001.
[42] Zhong G, Yan S, Huang K, Cai Y, Dong J. Reducing and stretching deep convolutional activation features for accurate image classification. Cogn Comput 2018;10:179–86. https://doi.org/10.1007/s12559-017-9515-z.
[43] Xu Y, Jia Z, Ai Y, Zhang F, Lai M, Chang EI. Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation. In: 2015 IEEE Int Conf Acoust Speech Signal Process; 2015. p. 947–51.
[44] Kulkarni A, Panditrao A. Classification of lung cancer stages on CT scan images using image processing. In: Proc 2014 IEEE Int Conf Adv Commun Control Comput Technol; 2015. p. 1384–8.

[45] Shashi B, Rana S. A review of medical image enhancement techniques for image processing. Int J Curr Eng Technol 2011;5:1282–6. https://doi.org/10.14741/Ijcet/22774106/5.2.2015.121.
[46] Building powerful image classification models using very little data. https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html. [Accessed 15 December 2018].
[47] Rokach L, Maimon O. Decision trees. https://doi.org/10.1007/0-387-25465-X, 2005.
[48] Hassanat ABA. Two-point-based binary search trees for accelerating big data classification using KNN. PLoS ONE 2018;13:e0207772.
[49] Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with different numbers of nearest neighbors. IEEE Trans Neural Netw Learn Syst 2018;29:1774–85. https://doi.org/10.1109/tnnls.2017.2673241.
[50] Wang H, Fan Y, Fang B, Dai S. Generalized linear discriminant analysis based on euclidean norm for gait recognition. Int J Mach Learn Cybern 2018;9:569–76. https://doi.org/10.1007/s13042-016-0540-0.
[51] Tharwat A, Gaber T, Ibrahim A, Hassanien AE. Linear discriminant analysis: a detailed tutorial. AI Commun 2017;30(2):169–90. https://doi.org/10.3233/aic-170729.
[52] Hasan AM. Linear regression based feature selection for microarray data classification. Int J Data Mining Bioinform 2015;11:167–79. https://doi.org/10.1504/ijdmb.2015.066776.
[53] Burges CJC. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 1998;2:121–67. https://doi.org/10.1023/A:1009715923555.
[54] Cömert Z, Kocamaz AF. Comparison of machine learning techniques for fetal heart rate classification. Acta Phys Pol A 2017;132:451–4. https://doi.org/10.12693/APhysPolA.131.451.
[55] Parikh KS, Shah TP. Support vector machine – a large margin classifier to diagnose skin illnesses. Proc Technol 2016;23:369–75. https://doi.org/10.1016/j.protcy.2016.03.039.
[56] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 2017;15:41–51. https://doi.org/10.21873/cgp.20063.
[57] Gao B, Pavel L. On the properties of the softmax function with application in game theory and On the properties of the softmax function with application in game theory and reinforcement learning; 2018.
[58] Ding C, Peng H. Minumum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 2005;03:185–205. https://doi.org/10.1142/S0219720005001004.
[59] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226–38. https://doi.org/10.1109/tpami.2005.159.
[60] Peng H. Python binding to mRMR feature selection algorithm. https://github.com/fbrundu/pymrmr.
[61] Park H, Lee JH, Oh Y, Ha S, Lee S. Training deep neural network in limited precision. arXiv:1810.05486, 2018.
[62] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts. arXiv:1608.03983, 2017.
[63] Reeskamp P. Is comparative advertising a trade mark issue? Eur Intellect Prop Rev 2008;30:130–7. https://doi.org/10.1145/2623330.2623612.
[64] Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation 2011;2:37–63. https://doi.org/10.9735/2229-3981.
[65] Chang AB, Ooi MH, Perera D, Grimwood K. Improving the diagnosis, management, and outcomes of children with pneumonia: where are the gaps? Front Pediatr 2013;1:29. https://doi.org/10.3389/fped.2013.00029.
[66] Rabhi S, Jakubowicz J, Metzger M-H. Deep learning versus conventional machine learning for detection of healthcare-associated infections in french clinical narratives. Methods Inf Med 2019;58(01):031. https://doi.org/10.1055/s-0039-1677692.
[67] Sriram A, Kalra S, Tizhoosh HR. Projectron – a shallow and interpretable network for classifying medical images. arXiv:1904.00740, 2019.
[68] Saraiva A, Ferreira N, Lopes de Sousa L, Costa N, Sousa J, Santos D, et al. Classification of images of childhood pneumonia using convolutional neural networks. In: Proceedings of bioimaging; 2019. p. 112–9.
[69] Livieris I, Kotsilieris T, Anagnostopoulos I, Tampakas V. DTCo: an ensemble SSL algorithm for X-rays classification. In: Advances in experimental medicine and biology. Springer; 2018. p. 1–10.
[70] Livieris I, Kanavos A, Tampakas V, Pintelas P. A weighted voting ensemble self-labeled algorithm for the detection of lung abnormalities from X-rays. Algorithms 2019;12:64. https://doi.org/10.3390/a12030064.
[71] Hooda N, Jakhar K. Big data deep learning framework using keras: a case study of pneumonia prediction. In: ICCCA; 2018.
[72] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172:1122–31. https://doi.org/10.1016/j.cell.2018.02.010.